



Universitat Autònoma
de Barcelona

TESI DOCTORAL:
**AVANCES EN LA APLICACIÓN DE LA
ESPECTROSCOPIA NIR EN LA INDUSTRIA
FARMACÉUTICA. INTRODUCCIÓN A PAT Y
TÉCNICAS DE IMAGEN**

MANEL BAUTISTA MERCADER

Director: Marcelo Blanco Romía

PROGRAMA DE DOCTORAT DE QUÍMICA
Departament de Química
Facultat de Ciències
2009

Memòria presentada per aspirar al Grau de Doctor per Manel Bautista Mercader

Manel Bautista Mercader

Vist i plau

Dr. Marcelo Blanco Romía
Catedràtic de Química Analítica

Bellaterra, 20 de Maig de 2009

AGRAÏMENTS

Aquesta tesi doctoral s'ha realitzat gràcies a les següents ajudes institucionals:

Ministerio de Educación y Ciencia, Madrid.

Projectes: “METODOLOGÍAS BASADAS EN ESPECTROMETRÍA NIR PARA EL CONTROL ANALÍTICO DE DE PROCESOS” (BQU2003-04247), i “LA ESPECTROSCOPIA NIR: UNA HERRAMIENTA PARA LA TECNOLOGIA ANALITICA DE PROCESOS (PAT)” (CTQ2006-12923).

EMPRESA/ADMINISTRACIÓN FINANCIADORA: Laboratorios Menarini S.A

Projectes: “Desarrollo de Métodos analíticos de multicomponentes basados en datos espectrofotométricos en el Infrarrojo Próximo (NIR)

Grup de Quimiometria Aplicada

Unitat de Química Analítica

Departament de Química

Universitat Autònoma de Barcelona

Edifici Cn, 08193 Bellaterra

Tot és psicològic

Poder aquestes son les paraules més complicades d'escriure, sobretot després de gairebé 10 anys en aquesta universitat dels quals he conviscut 3 i mig en aquest grup d'investigació.

Vull agrair sincerament al Doctor Marcelo Blanco per haver pensat en mi per a incorporar-me en el Grup de Quimiometria Aplicada als inicis de 2006, lloc en el qual m'he format realment com a l'investigador que soc a dia d'avui. Dins d'aquest grup he tingut la sort de compartir feina i hores amb companys que, a més de bons investigadors, han estat bons amics durant aquest viatge: Manel, Jordi, Anna, David, Chacho, Lisi, Ricardo, Juan, Ruben, Miguel, Mari, Miriam, Jose, Quim, Ester,....buff quanta gent. També voldria agrair als altres professors del grup d'Investigació per tot el suport durant aquest temps.

Voldria agrair a la gent que m'ha recolzat en la meua estada a Copenhaguen. Així, voldria donar les gràcies a Gilda, i als professors R. Bro i F Van der Berg. També per suposat a tots els estrangers amb els quals he compartit moments molt intensos fora de casa.

La meua família, clau en el meu creixement tant com a persona, com a investigador. Un apren de totes les experiències. Gràcies pel vostre suport.

Gracias Bea por todo.

Podria seguir escrivint, doncs he tingut la sort de rebre el suport per part de molta gent. Confio en que podré agrair personalment a la resta de gent.

Aquí s'acaba un viatge molt intens, a dia 26 de maig de 2009.

Gracies a tots

ABSTRACT

1. INTRODUCCIÓN

1.1 INDUSTRIA FARMACÉUTICA

1.2 ESPECTROSCOPIA NIR

1.2.1 INTRODUCCIÓN

1.2.2 FUNDAMENTOS DE LA TÉCNICA

1.2.3 INSTRUMENTACIÓN

1.3 APLICACIONES NIR EN LA INDUSTRIA FARMACÉUTICA

1.3.1 ESPECTROSCOPIA NIR CLÁSICA

1.3.2 ESPECTROSCOPIA NIR EN EL CONTROL DE PROCESOS

1.3.3 ESPECTROSCOPIA NIR IMAGEN

2 HERRAMIENTAS PARA EL ANÁLISIS DE DATOS. QUIMIOMETRÍA

2.1 MODELADO Y CONTROL

2.1.1 INFORMACIÓN CUALITATIVA

2.1.2 INFORMACIÓN CUANTITATIVA

2.2 HERRAMIENTAS PARA EL ANÁLISIS DE IMAGEN

2.3 OTROS

2.4 ADAPTACIÓN A TÉCNICAS DE IMAGEN

3. OBJETIVOS

4. METODOLOGÍA Y DISCUSIÓN DE LOS RESULTADOS

4.1 METODOLOGÍA EXPERIMENTAL

4.2 METODOLOGÍA INSTRUMENTAL.

4.2.1 INSTRUMENTACIÓN NIR

4.2.2 TÉCNICAS INSTRUMENTALES DE REFERENCIA

4.3 METODOLOGÍA DE TRATAMIENTO DE DATOS.

4.3.1 SOFTWARE

5. RESULTADOS Y DISCUSIÓN

5.1 DETERMINACIÓN DE PARÁMETROS DE INTERÉS EN PRODUCTO ACABADO

5.2 DETERMINACIÓN PARÁMETROS DURANTE PROCESO (*ON LINE MONITORING*)

5.2.1 INFORMACIÓN CUALITATIVA

5.2.2 INFORMACIÓN CUANTITATIVA

5.3 ESTUDIOS INICIALES NIR IMAGEN

5.3.1 ANÁLISIS CUALITATIVO

5.3.2 ANÁLISIS CUANTITATIVO

6. ARTICULOS ACEPTADOS PARA LA TESIS DOCTORAL

6.1 *PREPARING CALIBRATION SETS FOR USE IN PHARMACEUTICAL ANALYSIS BY NIR SPECTROSCOPY.*

6.2 *PHARMACEUTICAL GEL ANALYSIS BY NIR SPECTROSCOPY. DETERMINATION OF THE ACTIVE PRINCIPLE AND LOW CONCENTRATION OF PRESERVATIVES.*

6.3 *API DETERMINATION BY NIR SPECTROSCOPY ACROSS PHARMACEUTICAL PRODUCTION PROCESS*

6.4 *ON-LINE MONITORING OF GRANULATION PROCESS BY NIR SPECTROSCOPY*

6.5 *STUDY OF PHARMACEUTICAL SAMPLES BY NIR CHEMICAL IMAGE AND MULTIVARIATE ANALYSIS*

7. CONCLUSIONES

RESUMEN

Conseguir una elevada calidad en los productos es un reto cada día más importante en la industria moderna. En este sentido se están desarrollando normativas que deben cumplir los productos para llegar al mercado. Las exigencias de calidad que requiere la industria farmacéutica implican un amplio y exhaustivo control de múltiples parámetros relacionados con cualquier etapa en la obtención del producto.

La Espectroscopia NIR presenta una serie de ventajas respecto a otras técnicas: es no invasiva, no destructiva de la muestra, presenta una sistema de mediad y de obtención de resultados rápida, bajo coste debido a la ausencia de reactivos, etc, que han impulsado su aplicación en diferentes campos.

Se realiza un estudio de la aplicación de la Espectroscopia NIR a la determinación de parámetros de interés en la industria farmacéutica con el objeto de proponer nuevas metodologías más rápidas y eficaces que sean alternativas reales a los métodos convencionales.

Los estudios desarrollan una amplia variedad de métodos que cubren desde la determinación de principios activos en muestras de proceso o de producto acabado, hasta propiedades físicas de estas mismas muestras en diferentes matrices y niveles de concentración.

Los métodos desarrollados han sido validados siguiendo las guías ICH y EMEA que permiten su aplicación como métodos de calidad suficiente para el control analítico de un proceso de producción.

En consonancia con la metodología PAT, se ha realizado un estudio de la monitorización on-line de un proceso de granulación en una planta de producción mediante el uso de una sonda NIR no invasiva. Los resultados obtenidos invitan al optimismo de cara a desarrollar nuevas metodologías en el futuro y que puedan facilitar los recursos para obtener un producto final con la calidad deseada.

La memoria incluye también el primer trabajo desarrollado en el grupo de investigación en las técnicas de Imagen Química, y cuyos resultados e impacto se refleja en la publicación del trabajo en una revista de alto impacto científico.

ABSTRACT

Pharmaceutical production aims to provide a drug that meets strict quality conditions that ensure user satisfaction. This is a complex and varied procedure constituted by different phases. Quality assurance of final product can be achieved through a continuous control in the different stages, which require the application of different instrumental techniques capable of determining not only the chemical composition of the preparations but also for intermediate physical characteristics of them. These determinations consume considerable time reducing the productivity. This has led to the Food and Drug Administration (FDA) to promote an initiative, Process Analytical Technology (PAT), which aims to incorporate new technologies to the production process to provide a better and wider knowledge of the production process, reducing the time control and increase productivity.

A wide variety of instrumental techniques capable of providing solutions to this challenge have been proposed, but certainly the NIR spectroscopy is the most appropriate technique for their characteristics such as speed, no sample preparation, determination of composition and physical properties, ease incorporation of the productive process, and so on.

In order to improve various stages associated with pharmaceutical production processes to ensure the final quality of the drug and also intermediate stages of the process, we have approached the study of the application of NIR in some situations of interest to the manufacturing process through implementing systems to conventional NIR spectroscopy and imaging techniques Chemical latest deployment.

The objective of this thesis is to develop new applications of NIR spectroscopy that may be interesting in the production process in order to broaden awareness of it and that can improve the quality of the final product as the manufacturing process. The ultimate aim consists in applying these developed methodologies in the production process, either in control or in the intermediate stages of the finished product, which will be the demonstration that it has met its objectives.

1. INTRODUCCIÓN

1. INTRODUCCIÓN

1.1 INDUSTRIA FARMACÉUTICA

La industria farmacéutica es una actividad industrial altamente regulada por diferentes organismos tanto nacionales como supranacionales cuyos controles de calidad obligan a realizar un alto número de determinaciones de características tanto físicas como químicas de las materias primas, de los productos intermedios y también de los productos acabados. Este control exige la aplicación de un elevado número de metodologías analíticas que a menudo requieren protocolos de análisis de ejecución larga y tediosa y que impiden obtener resultados de un modo rápido. El principio activo (API) de un producto farmacéutico es generalmente determinado por cromatografía líquida de alta resolución (HPLC), que consume mucho tiempo, es una técnica destructiva y produce residuos. La determinación de propiedades físicas de las formulaciones implica el uso de una amplia variedad de otras técnicas instrumentales que a menudo son lentas e intensivas en mano de obra.

El control de calidad constituye un elemento esencial de los procesos de producción farmacéutica en virtud de los altos niveles de seguridad exigidos a partir de las formulaciones disponibles comercialmente. Garantizar la calidad de tales formulaciones implica someter los productos finales a una gran variedad de pruebas, que tienen un impacto negativo sobre la producción y la competitividad.

En respuesta a este problema, la FDA Orientación PAT [1] creó una nueva filosofía de trabajo destinada a la sustitución del control del producto final por la vigilancia durante todo el proceso de producción de parámetros de control con el fin de identificar cualquier debilidad y garantizar la calidad en el producto. Uno de los principales objetivos de esta iniciativa era mejorar la calidad de los productos finales y aumentar la eficiencia de los procesos de producción para conseguir de esta manera, construir la calidad en todo el proceso. Cuanto mejor es el conocimiento del proceso, mayor es la probabilidad de mejorar la calidad del producto. El cumplimiento de este objetivo

requiere un conocimiento profundo del efecto de algunas variables sobre el proceso y esos conocimientos pueden ser adquiridos por el seguimiento de la evolución de parámetros de calidad durante el propio proceso. La aplicación de la nueva filosofía de trabajo requiere el desarrollo y la utilización de técnicas analíticas rápidas y sencillas que permitan adquirir de forma rápida la información necesaria [2]. Las técnicas espectroscópicas son especialmente apropiadas para ser utilizadas en esta metodología, especialmente la espectroscopia de infrarrojo cercano (NIRS). De hecho, la técnica NIRS ha sido ampliamente utilizada con éxito en este contexto, como se refleja en el gran número de aplicaciones en la industria farmacéutica [3-6], donde su flexibilidad ha permitido la obtención de información analítica sin tratar de muestras de cualquier tipo con un mínimo de humanos intervención.

1.2 ESPECTROSCOPIA NIR

1.2.1 INTRODUCCIÓN

La región del infrarrojo está comprendida entre las longitudes de onda de 700 a 10^6 nm. Tanto por razones instrumentales como por las características de la interacción de la radiación con la materia, la región del infrarrojo se divide en tres zonas (tabla 1.1): infrarrojo cercano (NIR, Near Infrared), infrarrojo medio (MIR, Middle Infrared) e infrarrojo lejano (FIR, Far Infrared).

Región	Longitud de onda (nm)	Origen de la absorbancia
NIR	700-2500	Sobretonos y bandas de combinación de las vibraciones moleculares fundamentales.
MIR	2500-50000	Vibraciones moleculares fundamentales.
FIR	50000- 10^6	Rotaciones moleculares.

Tabla 1.1. División de la región del infrarrojo.

La radiación en la región del NIR fue descubierta en 1800 por William Herschel [7] siendo la primera región del espectro de absorción no visible descubierta. A pesar del temprano descubrimiento de esta radiación en la región del NIR, la aplicación de esta espectroscopia para la resolución de problemas en el ámbito de la química analítica fue lento, pasando casi 100 años, hasta 1880 cuando Abney y Festing registraron fotográficamente el espectro de líquidos orgánicos en la región espectral entre 700 y 1200 nm [8]. Posteriormente, durante la primera mitad del siglo XX, el interés de los espectroscopistas por la región infrarroja se centra en la región del infrarrojo medio, prestando poca atención a la zona vecina al visible, convencidos del poco interés de esta región para fines analíticos.

En la década de los 50, con la aparición de los primeros espectrofotómetros comerciales equipados con detectores fotoeléctricos capaces de registrar espectros en la zona del infrarrojo cercano, se desarrollaron las primeras aplicaciones en las que intervenía esta espectroscopia. En 1954 Wilbur Kaye[9] presentó su trabajo en el cual se registraban líquidos orgánicos entre 700 y 3500 nm asignando las diferentes bandas de absorción a los diferentes grupos funcionales.

La década de los 60 fue una etapa muy destacada en el desarrollo de esta técnica cuando un grupo del Ministerio de Agricultura de los EEUU, USDA (*United States Department of Agriculture*), liderado por Karl Norris empezó a utilizar esta metodología para el estudio de muestras en el ámbito agroalimentario [10-11]. A partir de ese momento, el interés por la espectroscopia NIR experimentó un cambio substancial. Hasta finales de los años 70 los instrumentos comerciales fueron principalmente instrumentos de filtros diseñados para aplicaciones específicas. Es a partir de aquí cuando se empiezan a desarrollar nuevos equipos mejorados y con más prestaciones y se construyen los primeros espectrofotómetros, que con la llegada de la era digital, los avances en el diseño y desarrollo de nueva instrumentación, con una electrónica más estable y con ópticas mejoradas, permitieron registrar espectros de manera rápida y altamente reproducible (instrumentos de barrido, interferómetros,

etc.) [12-13]. Igualmente importante fue el avance informático permitiendo desarrollar equipos informáticos cada vez más potentes y a precios razonables, clave para el progreso e implantación de la técnica.

Gracias a estos avances tecnológicos los espectroscopistas empezaron a mostrar interés por la técnica y el número de aplicaciones en la región NIR ha aumentado considerablemente en los últimos años. Sectores como el químico [14] petroquímico [15], agroalimentario [16] y farmacéutico [17] han utilizado la técnica cada vez más con resultados exitosos y siendo finalmente aceptada como método analítico, implementándose en algunos casos, bajo normas de garantía de calidad [18]. Además de los avances tecnológicos, las características intrínsecas de la técnica llamó la atención de muchos sectores que empezaron a desarrollar metodologías en esta región con la finalidad de sustituir ésta por otras técnicas.

Entre las principales ventajas de la Espectroscopía NIR, destacan: la posibilidad de registrar el espectro de muestras en estado sólido, líquido y gas, siendo el registro de los espectros muy rápido, sin necesidad de preparación de muestra ni destrucción de la misma al ser analizada. Los espectros NIR aportan información de propiedades químicas y físicas, por lo que varios parámetros pueden ser determinados en un mismo análisis. Dadas las ventajas de la espectroscopia NIR frente a otras técnicas instrumentales, puede considerarse como una técnica de relevancia dentro de la Tecnología Analítica de procesos.

El desarrollo de nuevas herramientas quimiométricas para el tratamiento de datos, la miniaturización de los componentes instrumentales que facilitan la portabilidad de los equipos para trabajar in-situ y la reciente combinación de la Espectroscopia NIR con técnicas de imagen, hacen que esta técnica esté en plena fase de auge y expansión[19-21]

1.2.2 FUNDAMENTOS DE LA TÉCNICA

Origen de la absorción

El origen de la absorción de las bandas en el Infrarrojo cercano es el mismo que para las bandas del IR medio: una molécula absorbe radiación electromagnética en esta región si la energía de la radiación corresponde a la diferencia energética entre 2 niveles vibracionales, además de producirse un cambio en el momento bipolar de la molécula. Cuando se dan estas dos circunstancias, ocurre una transferencia neta de energía que da lugar a un cambio en la amplitud de la vibración molecular y como consecuencia absorbe la radiación [22] El fenómeno espectroscópico que provoca la absorción de energía por parte de la materia es debido: a las rotaciones moleculares en el IR lejano, a bandas fundamentales de vibración en el IR medio, y finalmente a bandas de combinación de las bandas fundamentales y a sobretonos en el IR cercano. [23-24]

Cuando se trata de especies homonucleares como O₂, N₂ o Cl₂, el momento bipolar no se altera durante la vibración o la rotación y, en consecuencia, este tipo de compuestos no absorben radiación en el infrarrojo.

Los movimientos vibratorios que presenta una molécula pueden ser básicamente de dos tipos:

- Estiramiento. Se produce un cambio en la distancia interatómica a lo largo del eje del enlace de dos átomos durante la vibración.
- Flexión. Además de la modificación en la distancia interatómica se produce un cambio en el ángulo de enlace.

A diferencia del Infrarrojo medio, en la región del NIR no aparecen las bandas de vibración fundamentales ($\Delta U = \pm 1$). Las bandas debidas a los sobretonos corresponden a transiciones energéticas en las cuales $|\Delta U| > 1$, aunque siendo los

sobretonos con $\Delta U = \pm 2, 3$ y 4 , los únicos con probabilidad de ser observados en esta región. Los sobretonos aparecen a frecuencias ligeramente menores que las teóricas como múltiplos de las frecuencias fundamentales, debido al diferente espaciado entre los niveles energéticos.

Las bandas de combinación se producen exclusivamente en moléculas poliatómicas y son debidas a cambios simultáneos en la energía de dos o más modos de vibración. La interacción entre ambos modos da lugar a las bandas de combinación, cuya frecuencia es combinación lineal de las frecuencias fundamentales que interaccionan:

$$U_{\text{comb}} = n_1 U_1 \pm n_2 U_2 \pm n_3 U_3 \pm \dots \dots \dots (1)$$

Donde n_1, n_2, \dots son números enteros.

Las transiciones no fundamentales son menos probables que las transiciones entre niveles consecutivos. Por esta razón las bandas NIR son de menor intensidad que las que aparecen en la región del Infrarrojo medio. Además, son bandas más anchas y no están bien definidas debido al solapamiento de los sobretonos y bandas de combinación. [25].

Las bandas más frecuentes en la región del NIR son debidas a enlaces que contienen átomos con elevadas diferencias en el peso molecular (por ejemplo C-H, O-H, N-H,..), lo cual provoca un aumento en la anarmonicidad del enlace.

En la figura 1.1 se muestran las regiones NIR donde absorben los diferentes enlaces. Las bandas de los grupos C=O, C-C, C-F, C-Cl, en general son muy débiles o no aparecen en la región NIR. Las vibraciones fundamentales de estos grupos tienen bajas frecuencias en la región IR y por tanto, los primeros sobretonos también aparecen en esta región.

Una característica de la zona del infrarrojo cercano es que las interacciones entre moléculas, como puentes de hidrógeno, afectan al espectro NIR, por lo que un

espectro NIR también incorpora información sobre la estructura cristalina de la sustancia.

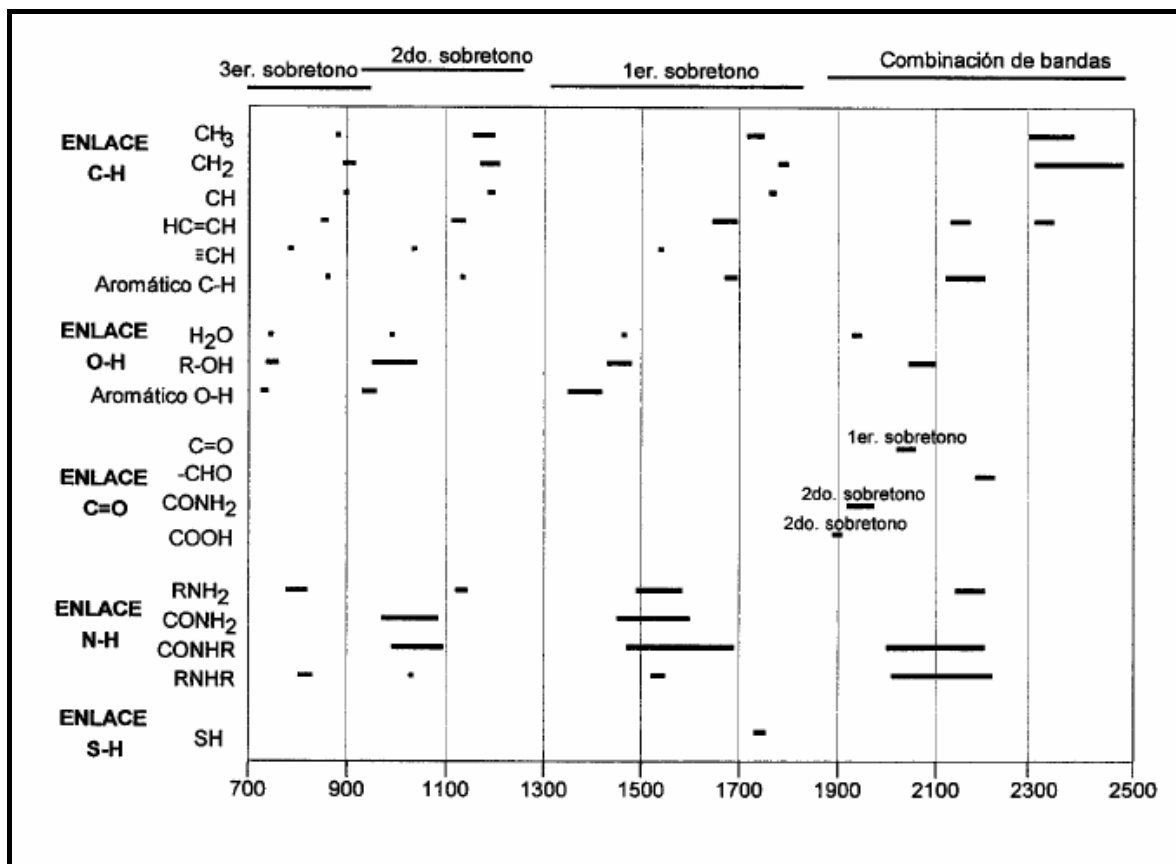


Figura 1.1. Representación gráfica de las zonas del espectro NIR donde se manifiestan los principales grupos funcionales.

Tipos de medida en la región NIR

Una de las grandes ventajas de la espectroscopia NIR es la versatilidad y adaptabilidad que presenta para analizar muestras de distinta naturaleza. Para las muestras en forma sólida, la medida efectuada es por reflectancia difusa. En cambio, muestras líquidas suelen medirse por transmisión de radiación. Un caso intermedio es el de muestras que se miden por transflectancia, en el que parte de la luz incidente es reflejada en la muestra y otra parte la atraviesa, para ser reflejada por un dispositivo transflectante, diseñado de tal forma, que también delimita el camino óptico. En la figura 1.2 se muestran los tres modos de interacción de la radiación NIR con la materia.

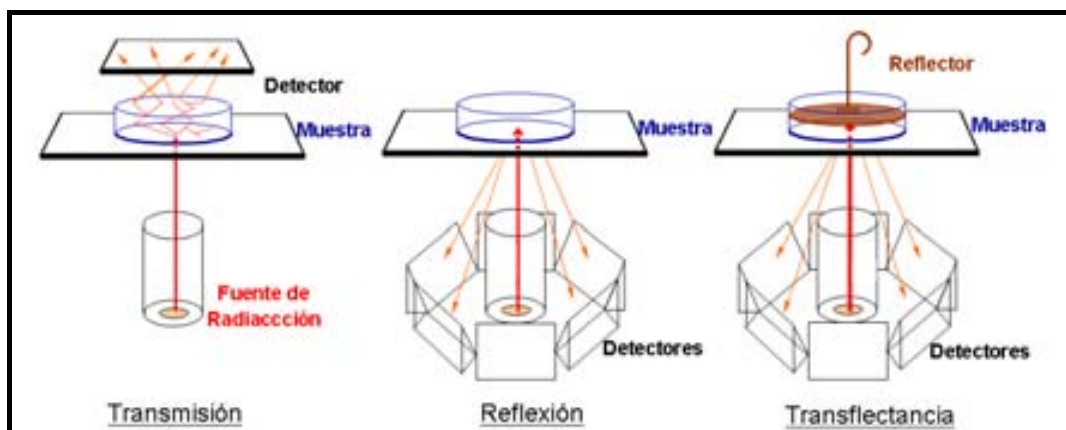


Figura 1.2. Diferentes modos de interacción de la materia con la radiación NIR. Imagen extraída de la Tesis Doctoral de Antonio Peinado (2005)

Medidas para reflectancia.

La espectroscopia de reflectancia estudia la radiación reflejada por una muestra que ha estado irradiada. Consta de dos componentes: La difusa y la especular. La componente especular descrita por las leyes de Fresnel, aporta poca o nula información acerca de la composición de la muestra. El detector se coloca en un ángulo determinado frente a la muestra de manera que se minimizan las radiaciones especulares.

La componente difusa se genera en todas las direcciones y predomina cuando la superficie irradiada es débilmente absorbente a la longitud de onda incidente y la penetración de la radiación es mayor en relación a la longitud de onda. La señal analítica medida se expresa como absorbancia aparente:

$$A_{ap} = \log \frac{1}{R} = a c \quad (2)$$

siendo **R** la reflectancia relativa ($R = R_{muestra} / R_{standard}$), **a** una constante de proporcionalidad y **c** la concentración.

Medida por Transmisión

La absorción de la radiación NIR sigue la ley de Lambert-Beer por lo que las medidas por transmisión pueden ser utilizadas con finalidad cuantitativa. La absorbancia puede ser definida:

$$A = \log \frac{I_0}{I} = \log \frac{I_0}{I} = \varepsilon b c \quad (3)$$

siendo **T** la transmisión, **I₀** la intensidad de radiación incidente, **I** la intensidad de la radiación transmitida, **ε** la absorptividad molar, **b** el camino óptico y **c** la concentración.

No obstante, de manera análoga a las regiones del visible y el MIR, es pueden producir desviaciones del comportamiento de la ley, debido a cambio en los enlaces por puentes de hidrógeno, formación de complejos, u otros procesos químicos.

Medida por transflectancia

La muestra, generalmente líquida o semi-líquida, se coloca en una cubeta junto con una superficie reflectante por una de las caras. El haz de radiación incidente entra por la cara transparente de la cubeta, atraviesa la muestra y se refleja en la otra cara de la cubeta. Así retorna de nuevo atravesando la muestra y la radiación es captada por el detector.

1.2.3 INSTRUMENTACIÓN

Debido a la baja intensidad de las bandas NIR, el nivel de exigencia de los espectrofotómetros NIR, en términos de nivel de ruido permisible y estabilidad instrumental, debe ser mayor que en otros instrumentos, sobre todo si se pretende aplicar al análisis cuantitativo. Ésta es una de las causas de la tardía aparición de los primeros espectrofotómetros NIR, que empezaron a comercializarse sólo cuando los

avances tecnológicos permitieron obtener niveles adecuados de sensibilidad, reproducibilidad y ruido.

El esquema básico de un instrumento NIR no difiere de cualquier espectrofotómetro. Sus componentes básicos son: fuente de radiación, sistema de selección de longitud de onda, compartimento de muestra y detector.

Fuente de radiación

La fuente de radiación más utilizada es la lámpara halógena de filamento de tungsteno con ventana de cuarzo, capaz de proporcionar un espectro continuo en la región de 320-2500nm. Otras fuentes de radiación que pueden utilizarse son los denominados LEDs (*Light Emitting Diodes*) [26-27], que dependiendo de su composición pueden llegar a emitir hasta los 1600 nm, pero no por encima de esa longitud. Los instrumentos que incorporan este último sistema de fuente de radiación no requieren un sistema de selección de longitudes de onda.

Sistemas de selección de longitudes de onda

Es el componente esencial que permite obtener la información espectral a cada longitud de onda. El selector debe proporcionar un ancho de banda estrecho respecto al ancho de banda que está midiendo y debe ser preciso y exacto para la longitud de onda analítica.

Los equipos pueden clasificarse básicamente en dos tipos: dispersivos y no dispersivos. Dentro de los sistemas dispersivos encontramos los monocromadores. Los monocromadores son dispositivos que descomponen la luz policromática que proviene de la fuente de radiación en longitudes de onda discretas. La radiación entra en forma de haz estrecho y un elemento dispersante, que puede ser un prisma o una red de difracción, la descompone. Los más utilizados actualmente son los que incorporan una red de difracción, ya que son más baratos, proporcionan mejor separación de longitudes de onda y dispersan linealmente la radiación [28].

El conjunto de sistemas no dispersivos es más amplio debido a su mayor utilización actualmente. Dispone de equipos con filtros adicionales, con filtros optoacústicos (AOTF) e instrumentos de transformada de Fourier (FT), ambos con características bastante diferentes entre sí. La selección de longitudes de onda mediante filtros se realiza interponiendo materiales específicos entre la muestra y la fuente de radiación, permitiendo el paso de longitudes de onda selectivas.

Compartimiento de muestra

Los instrumentos NIR permiten registrar el espectro tanto de muestras sólidas, líquidas y gaseosas. La ausencia de pretratamiento de muestra para el registro de su espectro, permite disponer de gran cantidad de accesorios adaptables a cada situación. Los 3 tipos de medida ya se mostraron en la figura 1.2 donde ya se puede observar las diferentes disposiciones en función del tipo de muestra.

Detector

Es el dispositivo encargado de recibir la señal luminosa que proviene de la muestra y la transforma en señal eléctrica. Detectores habituales en espectroscopia en el infrarrojo próximo son los construidos con semiconductores (como InGaAs, PbS, InAs, InSb, Si,...). Hasta 1100 nm el semiconductor más utilizado en los detectores es el Si, y en la región entre 1100 y 2500 nm el material fotoconductor más utilizado es el PbS.

1.3 APLICACIONES NIR EN LA INDUSTRIA FARMACÉUTICA

El control de calidad farmacéutica requiere metodologías que permitan obtener la información de interés de una forma rápida y sencilla, sin necesidad de realizar una preparación de la muestra.

La Espectroscopia NIR cumple con los requisitos exigidos, como demuestra el elevado número de aplicaciones desarrolladas para su utilización en la industria farmacéutica [29-31]. Su gran capacidad de adaptación permite analizar casi cualquier tipo de muestra con una mínima intervención del analista.

Se ha realizado una clasificación dentro de las aplicaciones en la industria farmacéutica con la finalidad de observar un progreso de las posibilidades prácticas de esta técnica, que ha ido evolucionando los últimos años, aportando solución a distintos problemas, desde la monitorización de los procesos químicos, hasta las técnicas de Imagen química, que abren la posibilidad de plantearnos nuevos desafíos y que permitirán obtener un mayor y mejor conocimiento dentro del análisis en la industria en general, y farmacéutica en particular.

En la presente memoria puede diferenciarse la aplicación en la industria farmacéutica en función de los objetivos y en función de la información obtenida. Por ello se han clasificado en 3 apartados: Espectroscopia NIR, Espectroscopia NIR para el Control de Procesos, y técnicas espectroscópicas NIR Imagen.

1.3.1 ESPECTROSCOPIA NIR

La principal aplicación en la industria farmacéutica consiste en el análisis de distintos parámetros de interés en el producto final, o bien al final de las distintas etapas de producción.

Si bien, muestras extraídas de un proceso químico pueden extraerse y analizarse mediante esta metodología, el hecho de tener que llevar la muestra al instrumento, hace poco adecuado el uso de esta metodología de la espectroscopía NIR para el seguimiento de un proceso (siguiente apartado). A continuación se describen los parámetros relevantes:

Identificación

La identificación mediante Espectroscopia NIR del API, excipientes utilizados para la fabricación de productos, y los productos finales entre los mismos es necesaria para las pruebas de rutina de control de productos farmacéuticos procesos de fabricación.

La Espectroscopia NIR ha sido utilizada para controlar todos los pasos en el proceso de fabricación de drogas químicas incluidas síntesis, el análisis de las materias primas, productos intermedios proceso (después de mezclar, moler, secado, compactación) y preparados, y, finalmente, en el control de calidad. Los productos químicos pueden ser caracterizados además por las propiedades físicas como el tamaño de partícula, dureza o viscosidad.

La construcción de bibliotecas de espectros evita la necesidad de desarrollar un método específico para cada producto y permite su identificación en un sencillo, directo y rápido [32]. Todo el proceso puede ser automatizado, controlado por ordenador y realizado por operadores sin un elevado grado de conocimiento de la técnica.

El alto grado de solapamiento de bandas espectrales así como la ausencia de bandas estrechas dificulta la construcción de bibliotecas, por lo que el uso de herramientas quimiométricas permite extraer información relevante.

Determinación del contenido de API

El API corresponde a la sustancia activa presente en un fármaco y con propiedades terapéuticas. La actividad de un principio activo varía debido a la naturaleza de estos, pero siempre está relacionado a la cantidad ingerida o absorbida. Una forma de dosificación de un fármaco tradicionalmente se compone de dos componentes: el API, que es la droga en sí, y uno o varios excipientes cuya mezcla corresponde al placebo, inerte terapéuticamente.

Certificar la correcta dosificación de API en el producto final supone uno de los parámetros fundamentales para el control de calidad del producto. Sin embargo es necesario distinguir entre dos tipos de determinación:

Uniformidad de dosis: determina el contenido de API correspondiente a una dosificación. Por ejemplo, se determinó el contenido de API en un comprimido, o en un sobre.

Uniformidad de contenido: determina la concentración de API en un lote entero de producto. Para llevar a cabo esta operación es necesario obtener la uniformidad de dosis para varias muestras de un mismo lote.

Determinación de la humedad

El agua es uno de los parámetros que presentan mayor señal en la región del infrarrojo cercano: dos bandas anchas y de gran intensidad aproximadamente en 1490 y 1920 nm correspondientes al primer sobretono y a la banda de combinación de las bandas fundamentales de vibración respectivamente. Como consecuencia de la elevada intensidad del agua en la región NIR, su determinación ha sido ampliamente utilizada desde los inicios de la técnica [33-35].

Durante distintas etapas del proceso de producción de productos farmacéuticos sólidos, como la granulación o el secado, la determinación de este parámetro resulta esencial para continuar el proceso productivo. Sin embargo, el análisis de productos que contiene principalmente agua en sus matrices, dificulta la determinación de otros parámetros debido a su elevada intensidad en la señal, interfiriendo en la señal de otros analitos de interés con menor intensidad en la región NIR.

Determinación del tamaño de partícula

El tamaño de partícula de una muestra sólida influye directamente sobre el efecto de *scattering* en medidas NIR por reflectancia difusa [36-37]. Este efecto produce desplazamientos de línea base que permite la determinación del tamaño medio de partícula de muestras sólidas. El análisis de este parámetro, de la misma forma que el agua, constituye un paso relevante para continuar en muchas etapas de los procesos de fabricación de fármacos.

Determinación de Densidad

La densidad se define como el cociente entre *la* masa de un cuerpo *y el* volumen que ocupa. La densidad de un sólido farmacéutico depende de la naturaleza de la materia prima, además del tamaño de partícula. La determinación de este parámetro es de vital importancia, junto con la determinación de humedad y tamaño de partícula, cuando el proceso de fabricación final termina en una etapa de compactación.

Transformación polimórfica

El polimorfismo puede definirse como la capacidad que tiene una sustancia de existir como dos o más formas cristalina que tienen conformaciones distintas de las moléculas en la red cristalina [38].

Dentro de los productos farmacéuticos, existen algunos que poseen esta propiedad. Sin embargo no todas las conformaciones de una misma sustancia tienen las mismas propiedades (velocidad de disolución, estabilidad química o fluidez) que pueden tener importancia terapéutica, por lo que la determinación de la forma cristalina del producto farmacéutico constituye un gran interés en el control de calidad final.

Las técnicas espectroscópicas basadas en las transiciones vibracionales, rotacionales y de spin nuclear, permiten la identificación y cuantificación de las formas polimórficas.

Un caso especial es el la forma amorfa al carecer de estructura cristalina definida y se consideran pseudo-polimorfos. Varios estudios sobre la conversión entre la forma polimórfica y amorfa han sido determinados mediante la Espectroscopia NIR [39-40]

1.3.2 ESPECTROSCOPIA NIR EN EL CONTROL DE PROCESOS

La evolución de la espectroscopia NIR y de las técnicas espectrofotométricas en los últimos años, ha permitido ampliar su uso en múltiples procesos de la industria química. La evolución desde el concepto *off-line* hasta el *non-invasivo* ha conducido a “llevar la luz a la muestra” en vez de “llevar la muestra a la luz” [41]. La automatización

de la toma de muestra [42-43] o el uso de fibras ópticas introducidas en la línea de proceso [44-45], eliminan por completo la manipulación de la muestra, obteniéndose información de forma aun más rápida. Existen diversos diseños de fibras ópticas para adaptarse a la muestra, según el tipo de medida que sea necesaria, transmisión, reflectancia o transreflectancia.

Desde hace unos años, la Food and Drug Administration (FDA) puso en marcha la iniciativa PAT (Process Analytical Technologies) [1] que incentiva a la industria farmacéutica a utilizar nuevas metodologías de trabajo con la finalidad de diseñar, analizar y controlar todos los pasos de un proceso de fabricación. Este término, PAT, tiene además el valor añadido de conocer y asegurar la calidad del producto de interés en cualquier punto del proceso productivo.

Para llevar a cabo estas metodologías son necesarias técnicas que permitan realizar análisis en un intervalo pequeño de tiempo, además de poder trabajar en continuo sin necesidad de personal altamente especializado.

La Espectroscopia NIR cumple con estos requisitos: i) muestreo no destructivo y no invasivo, ii) monitorización en continuo con obtención rápida de resultados, iii) determinación on-line del punto final del proceso y de la calidad del producto final, por lo que se presenta como una técnica útil para desarrollar metodologías PAT. Además, ha sido recientemente aplicada a metodologías PAT para procesos de granulación de productos farmacéuticos [46-48].

La utilización de PAT en la industria farmacéutica ofrece muchas ventajas frente a los métodos tradicionales de control de calidad ya que no es necesario esperar al producto acabado para certificar su calidad, si no que ante cualquier alarma durante el proceso, existe la posibilidad de modificación y corrección, con la consiguiente mejora en eficiencia y productividad global de la empresa. Si bien la puesta a punto de una metodología PAT para un proceso de producción real puede resultar costosa, los resultados obtenidos pueden aportar un alto grado de conocimiento ayudando a la eficiencia del propio proceso.

Procesos de Granulación Húmeda

En la industria farmacéutica, la granulación es uno de los procesos más comunes durante la fabricación de un producto farmacéutico y cuya necesidad viene determinada por diferentes causas: i) generalmente la materia prima presenta en polvo con un tamaño de partícula muy fino por lo que dificulta su manipulación y transporte ya que el polvo queda adherido en diferentes partes de la maquinaria utilizada, ii) un elevado porcentaje de los fármacos sólidos suelen finalizar su fabricación con un proceso de compactación para producir comprimidos y iii) aumenta la uniformidad de la distribución del fármaco al disminuir la segregación. Con el objeto de mejorar el proceso se aplica una granulación apropiada que tiene por objeto aumentar el tamaño de partícula para facilitar la manipulación, transporte, dosificación del sólido y obtener un producto final con las características adecuadas de calidad [49,50].

Los procesos de granulación más frecuentes son los que utilizan *high shear mixer* y el *fluid bed system* (granulación húmeda), que son procesos en los que intervienen un número elevado de variables y que requieren de personal especializado para obtener el producto con las características adecuadas.

La granulación húmeda implica agregar una solución ligante a una mezcla sólida a fin de facilitar la agregación de partículas. La solución ligante de granulación es típicamente una mezcla de agua y un aglutinante. Una vez finalizada la etapa de adición tiene lugar la etapa de secado, que consiste en elevar la temperatura del flujo de inyección de aire con el objetivo de eliminar la humedad del medio. En este sentido, los gránulos empiezan a secarse y la colisión de los mismos conlleva a la fractura de estos gránulos en otros con un tamaño de partícula inferior. En función del producto final, las condiciones de las distintas etapas de granulación pueden variar.

El diseño, seguimiento y control de un proceso de granulación húmeda requiere un buen conocimiento de las variables asociadas al proceso y su producto.

Tradicionalmente, los procesos de granulación, se han controlado utilizando procedimientos visuales y manuales, y caracterizando los gránulos en términos de tamaño de partícula, contenido de humedad y densidad. Algunos métodos de referencia utilizados para determinar estas propiedades requieren utilizar una cantidad sustancial de granulado en muchos pasos del proceso (por ejemplo, durante la adición de la solución ligante y del secado del producto). A pesar de que proporcionan resultados aceptables, estas técnicas requieren mucho tiempo, además de analistas y técnicos experimentados, lo que excluye su utilización en la actualidad cada vez más las estrategias de expansión PAT.

La espectroscopia NIR se ha convertido en una herramienta útil para la rápida extracción. El objetivo principal de PAT, QbD (*Quality by Design*), es conseguir un mejor entendimiento y conocimiento sobre los procesos de producción.

Realizar ensayos en un equipo de planta piloto permite optimizar las condiciones de funcionamiento con anterioridad al proceso industrial, con miras a su transferencia a la planta de producción. Estudiar un proceso de granulación en planta resulta prácticamente inviable debido al alto costo de la optimización de los procesos normalmente de cientos de kilogramos de materia prima por lote. Sin embargo, el estudio de un proceso de granulación en escala permite extraer información valiosa para el entendimiento de dicho proceso, durante el cual existe variabilidad propia de los parámetros de interés en el proceso, y que por lo tanto pueden ser modelados.

El desarrollo de metodologías de análisis espectroscópicos NIR basado en la información pertinente a los procesos realizados bajo condiciones normales de funcionamiento puede a veces ser suficiente para mejorar el conocimiento sobre el destino de productos y procesos, acortar los tiempos de tramitación en cada paso y obtener un beneficio económico.

Repasando la bibliografía, varios estudios han examinado ya la determinación del tamaño de partícula y su distribución en lotes por espectroscopia NIR [51,52], utilizando las mediciones fuera de línea para facilitar la aplicación de los instrumentos

analíticos. Rantanen et al. [53] realizó un estudio en línea de granulación en una planta piloto para controlar los cambios en el tamaño de las partículas y contenido de humedad mediante el tratamiento matemático de los datos. Recientemente [54], se han propuesto varios modelos para la cuantificación de la humedad y el tamaño medio de partículas durante la granulación de un solo componente (lactosa) en una planta piloto. El control y optimización de variables potencialmente influyentes en el resultado de un proceso de granulación (por ejemplo, la temperatura del aire de entrada, caudal de carpetas, pulverizador de posición, las concentraciones de componentes) puede ser realizado en la planta piloto. Rambali et al. [55] han estudiado los cambios en el tamaño de grano en pequeña, mediana y gran escala. Otros autores [56-57] han desarrollado diseños experimentales eficaces para la determinación de parámetros físicos y químicos durante la granulación con el objetivo de optimizar las propiedades del producto final.

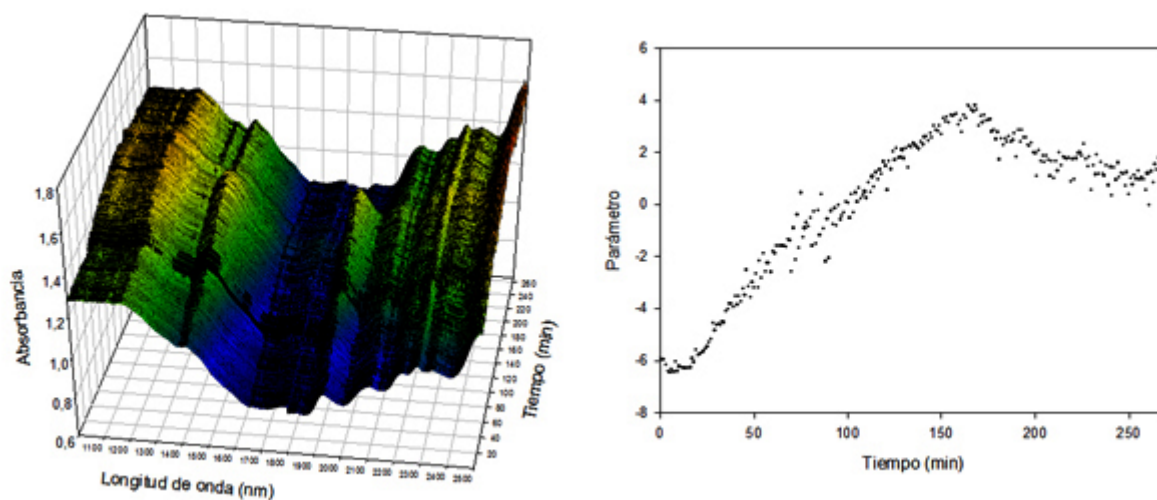


Figura 1.3. Representación de todos los espectros recogidos durante un proceso de granulación. La figura de la derecha muestra la evolución de un parámetro de interés a lo largo del proceso

Dentro de las aplicaciones de la Espectroscopia NIR para el control de procesos, además de la determinación puntual de los parámetros que se han comentado en el apartado de Espectroscopia NIR convencional, el mayor interés se centra en conseguir

gráficos o cartas de control, que permiten observar la evolución de un parámetro a lo largo de todo el proceso. Los gráficos de control ayudan en la detección de modelos no naturales de variación en los datos que resultan de procesos repetitivos, aportando criterios para detectar una falta de control estadístico. Un proceso se encuentra bajo control estadístico cuando la variabilidad se debe sólo a causas comunes.

1.3.3 ESPECTROSCOPIA NIR IMAGEN

Una imagen puede ser definida como la representación gráfica de un objeto real. En función del número y tipo de canales utilizados para obtener una imagen pueden obtenerse imágenes en escala de grises, imágenes RGB etc.

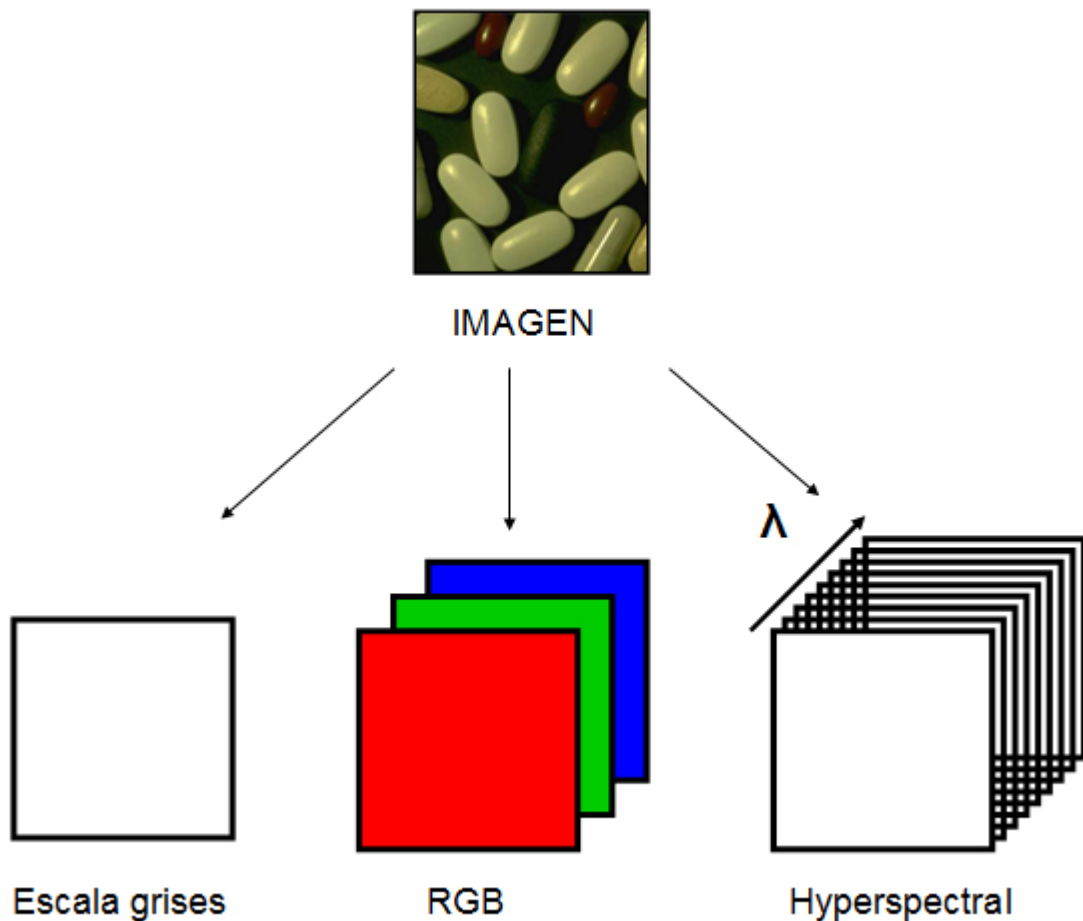


Figura 1.4. Representación de diferentes clases de obtención de imágenes.

Una imagen química puede ser definida como la representación de una imagen de la que se obtiene información química, como es el caso de la técnica de NIR imagen en la que se toman fotografías de un objeto a diferentes longitudes de onda, obteniendo valores de absorbancia en todos los puntos de la imagen. La fracción más pequeña de la imagen que contiene información se denomina píxel, y su tamaño depende de la resolución de la cámara y de la forma de registro de las imágenes.

Ampliando tanto como se desee o sea posible la cantidad canales, se obtienen imágenes hiperespectrales, que no son más que una imagen multivariante formada por tantas intensidades como intervalos de longitud de onda la formen.

El uso de imágenes químicas es un nuevo enfoque analítico que permite la identificación cuantitativa y espacial de las especies químicas en una muestra. En términos muy sencillos se puede considerar como un mapa de un producto químico. Mediante la combinación de espectroscopia infrarroja tradicional con las propiedades de imágenes microscópicas y macroscópicas, una sola medición rápida permite a los analistas obtener información espacial y espectral con gran facilidad sin estudios previos. La posibilidad de registrar un espectro en un amplio intervalo de longitudes de onda para cada píxel de la imagen de una muestra está enfocada al desarrollo de nuevas metodologías más robustas y fiables para extraer la información más relevante de los datos obtenidos.

Es bien sabido que la distribución espacial de los componentes en materiales complejos influye fuertemente en las propiedades físicas y químicas, siendo de gran importancia en la mayoría de procesos farmacéuticos donde uno de los objetivos principales es conseguir una correcta distribución de todos los analitos en los lotes de producción. Las técnicas de Imagen Química han encontrado muchos usos en la industria farmacéutica, las ciencias de polímeros, ciencia de materiales, medicina forense, patología, clínica y diagnóstico, así como la calidad industrial y control del proceso.

La utilización de las técnicas de NIR Imagen Química como metodología de trabajo en análisis farmacéutico esta en expansión como lo demuestran el creciente número de

trabajos que aparecen en la bibliografía [58-62]. Esta técnica permite la adquisición de información espectral y espacial de una muestra en un período corto de tiempo, una sola imagen contiene miles de espectros con lo que supera ampliamente la capacidad del NIR convencional. 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 [63].

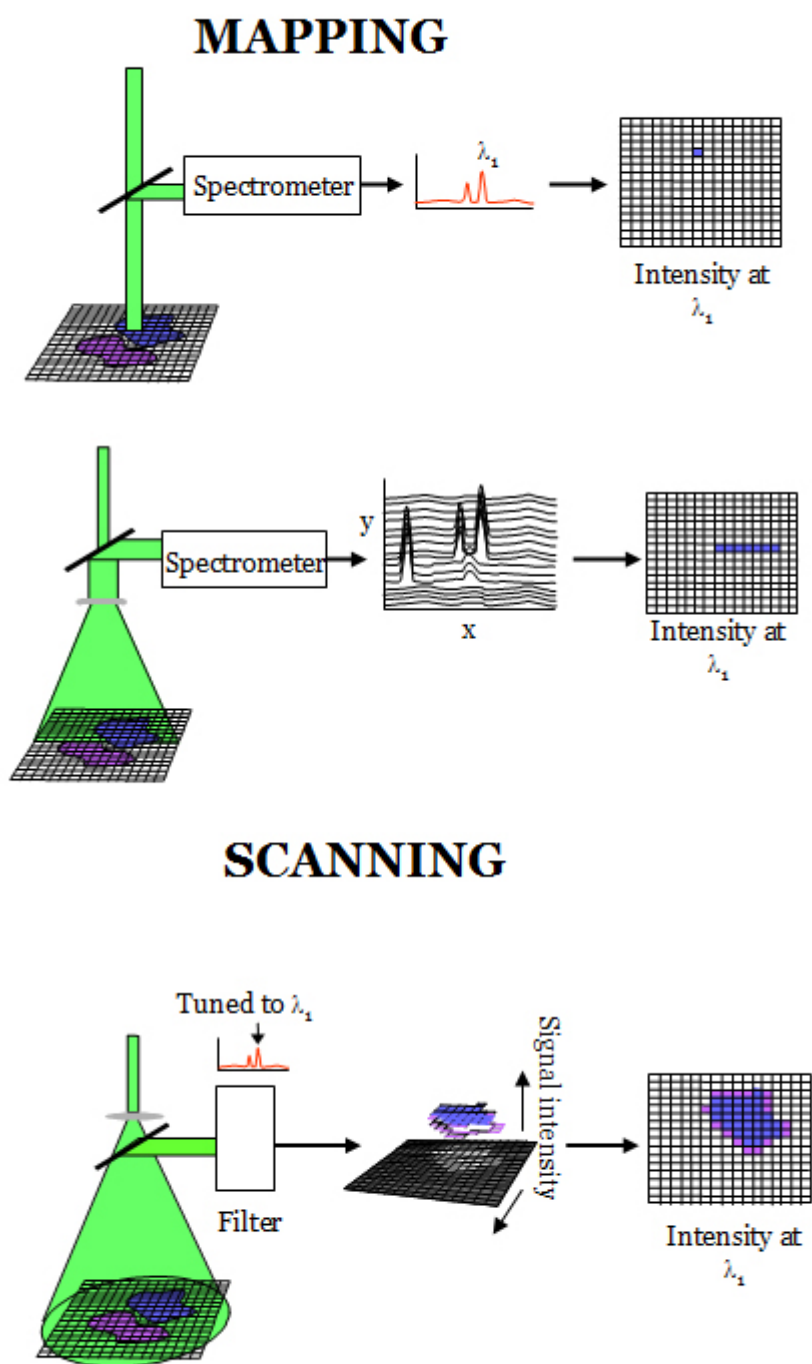


Figura 1.5. Clasificación de los diferentes métodos de registro de imagen.

Los dos tipos de registro de imagen más común en las técnicas de imagen NIR se muestran en la figura 2. El *mapping* está basado en crear mapas espectrales, a partir de la emisión de luz en un punto en concreto de la imagen como se muestra en el ejemplo de arriba, donde emite a una determinada longitud de onda en un punto. Así se toman espectros en todos los puntos de la muestra. Estas técnicas tienen una muy buena resolución espacial, sin embargo el tiempo requerido puede ser muy largo.

Dentro del *mapping* pueden distinguirse dos modos de registro:

El *Wisherbroom* mide el espectro de la imagen píxel por píxel, mientras que el *pushbroom* registra los espectros por líneas de píxeles, obteniendo la imagen en menor tiempo que el primer tipo de registro, aunque reduciendo la calidad de la imagen.

Otro tipo de registro de imágenes es el que se muestra en la parte inferior de la figura 1.5, que es el *scanning*, en el que se toman imágenes de toda la muestra a diferentes longitudes de onda, con lo que en tiempo de análisis se reduce muy considerablemente, aunque esto conlleva una disminución de resolución espacial.

La utilización de una u otra técnica depende de la información que se quiera obtener de la muestra a analizar. La principal ventaja de la técnica NIR Imagen con la Espectroscopia NIR convencional reside en la obtención de información en todos los puntos de la imagen, lo que permite realizar un estudio de la distribución y también el análisis cuantitativo de los componentes en la imagen.

La combinación de la técnica de NIR-Imagen, dentro del control de procesos, se presenta como un gran y novedoso avance en el desarrollo de metodologías PAT, combinando el potencial que presenta obtener información espectral y espacial, durante todo un proceso de fabricación de un preparado farmacéutico.

**2. HERRAMIENTAS
PARA EL ANÁLISIS
DE DATOS.
QUIMIOMETRÍA**

2 HERRAMIENTAS PARA EL ANÁLISIS DE DATOS. QUIMIOMETRÍA

Las técnicas analíticas utilizadas en la presente memoria proporcionan señales que contienen información de la composición y de otras propiedades de la muestra. Sin embargo, estas señales no son debidas a una única especie, por ejemplo, el espectro de una muestra será la suma de la contribución a la absorbancia de cada uno de los componentes que la forman a cada longitud de onda. Además, la elevada cantidad de información contenida en los espectros NIR requiere del uso de herramientas matemáticas para el tratamiento de los datos. En este sentido la quimiometría permite extraer la información presente en los espectros de un conjunto de muestras, ya sea de forma cualitativa como cuantitativa.

Svante Wold empleó la palabra “chemometrics” en 1972 para describir la disciplina de extraer información química relevante partiendo de un sistema químico experimental [63]

Sin embargo, Massart et al. presentó en el 1997 la siguiente definición:

“La quimiometría es la parte de la química que se sirve de las matemáticas, estadística y lógica formal para: diseñar o seleccionar procedimientos experimentales óptimos; proporcionar información química relevante a partir del análisis de señales analíticas y, finalmente, adquirir conocimiento de los sistemas químicos”.

La Quimiometría engloba herramientas, como el pretratamiento de señales (aplicados a datos experimentales para mejorar la calidad de la señal), métodos de reconocimiento de pautas [64] (técnicas multivariantes mediante las cuales se pueden establecer agrupaciones de muestras en función de su similitud y métodos de clasificación de nuevas muestras) y métodos de calibración [65,66] (técnicas multivariantes mediante las cuales se busca una relación entre la señal analítica y alguna propiedad de la muestra). Son muchas las técnicas quimiométricas que se pueden utilizar para llevar a cabo estos objetivos.

Tabla 2.1 Características principales de las herramientas quimiométricas utilizadas en la presente memoria.

HERRAMIENTA	OBJETIVOS	NECESIDAD CALIBRADO	BENEFICIOS	DESVENTAJAS
PCA	Selección de muestras. Modelado y control de procesos.	NO	Información directa. Muy útil en selección de muestras, previa calibración.	Solo información cualitativa. Difícil interpretación.
MCR-ALS	Modelado y control de procesos. Identificación. Determinación Cuantitativa.	NO (pero sí es conveniente información previa)	Obtención de información cualitativa y cuantitativa de forma sencilla.	Ambigüedades. Selectividad.
CLS	Modelado y control de procesos. Determinación Cuantitativa	NO	Método directo de cuantificación.	Requiere espectros puros. Variables físicas perjudican los resultados
PLS	Modelado y control de procesos. Determinación Cuantitativa	SI	Modelos de cuantificación robustos	Obtener calibrado robusto. Método de referencia
FCM-CLUSTERING	Clasificación Segmentación de imagen	--	No realiza clasificación absoluta, sino que otorga un grado de pertenencia de cada pixel a un grupo determinado.	Requiere saber el número de clusters previamente al análisis.
K-MEAN CLUSTERING	Clasificación Segmentación de imagen	--	Rápido y algoritmo de fácil implementación.	Requiere saber el número de clusters previamente al análisis.
COEFICIENTES CORRELACIÓN	Clasificación	--	Cálculo sencillo	Requiere espectros puros. Alta correlación espectral en la región NIR.

La Quimiometría ha sido utilizada en esta memoria para desarrollar modelos empíricos y aplicarlos para el análisis de muestras obtenidas en procesos de la industria farmacéutica [67-68]. Existen numerosos textos y publicaciones que contienen información extensa sobre la teoría y práctica de la Quimiometría en la industria

farmacéutica. Por ello en esta memoria únicamente se han recogido aquellos puntos más significativos que han sido utilizados para el desarrollo de los métodos propuestos. En la tabla 2.1 se muestra una breve descripción las herramientas utilizadas, en la presente Tesis. A partir del tipo de información que puede ser extraída de muestras farmacéuticas se han clasificado las herramientas.

Todas estas herramientas descritas en los siguientes apartados pueden ser aplicadas a las técnicas de Imagen, previo desdoblamiento del cubo hyperespectral en una matriz bidimensional.

2.1 MODELADO Y CONTROL

El proceso de desarrollo y evaluación de un modelo quimiométrico, ya sea destinado al análisis cualitativo o cuantitativo, consta de una serie de etapas básicas. De modo general, estas etapas tienen una serie de características comunes. A continuación se describen más ampliamente cada una de ellas [69].

Selección del conjunto de calibración

Las muestras que van a ser utilizadas para el desarrollo del modelo deben ser representativas de toda la variabilidad que pueda darse durante el proceso, con la finalidad de obtener métodos analíticos robustos. Según el modelo que vaya a ser desarrollado, la variabilidad que las muestras deben cubrir es distinta. Así pues, las muestras a utilizar en modelos cualitativos deben recoger la variabilidad que pueda darse en el tamaño de partícula, proveedor, etc. En los modelos cuantitativos, las muestras quizás deban contener esta variabilidad además de cubrir el rango del parámetro a determinar, por ejemplo la concentración del analito.

Métodos de referencia

El desarrollo de modelos cuantitativos requiere el conocimiento previo de las variables que van a ser determinadas. Para ello se utilizan los métodos de referencia que deben

proporcionar valores precisos y exactos, ya que de ello dependerá la calidad del modelo a desarrollar. En el apartado de metodología se describen los métodos de referencia utilizados en la presente memoria.

Pretratamientos espectrales

Los pretratamientos espectrales son correcciones del espectro basadas en tratamientos matemáticos con la finalidad de aumentar la calidad de la señal instrumental, reduciendo al máximo la información no asociada a la propiedad estudiada, como pueden ser desplazamientos de la línea base, ruido espectral, etc.

Se denomina ruido al componente espectral no relacionado con ninguna propiedad física o química de la muestra como puede ser el ruido instrumental o de iluminación. Además, otras fuentes de variación como la temperatura y humedad pueden aportar interferencias a la señal analítica. A continuación se presentan los pretratamientos más habituales utilizados en este trabajo.

- Suavizado de espectros (Smoothing Average): consiste en calcular el promedio de X valores consecutivos (tamaño de ventana) de Absorbancia, y colocarlo en la posición del primer valor. A continuación la ventana de tamaño X se mueve una posición en el espectro y de nuevo se calcula el promedio colocándolo en la segunda posición, sucesivamente. El suavizado tiene como objetivo eliminar ruido espectral que dificulta la extracción de la información relevante.

Sin embargo, el incorrecto uso del tamaño de ventana podría eliminar información relevante, por lo que es necesario optimizar el tamaño de ventana. En la figura 2.1 se observa en verde la eliminación de ruido espectral mientras que en la región de color rojo se ha eliminado información relevante.

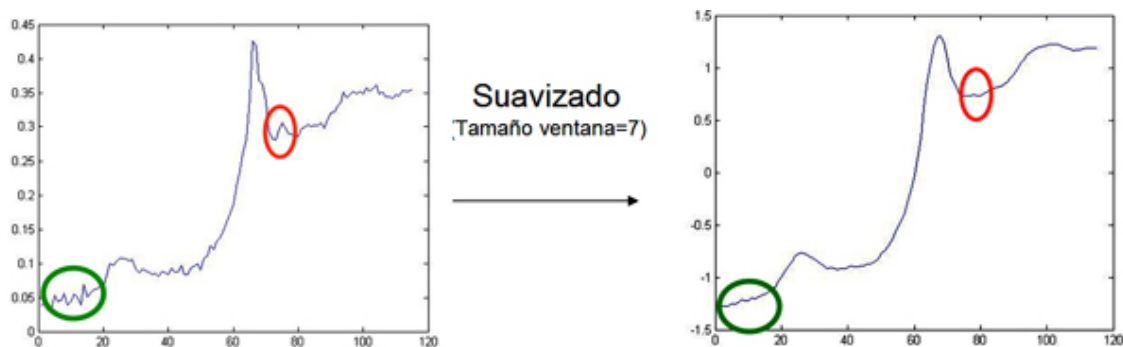


Figura 2.1. Ejemplo de suavizado de un espectro.

- SNV (Standard Normal Variate) [70]: corrige efectos de *scattering* provocados por diferencias físicas entre muestras. El método se aplica individualmente a cada espectro, y se obtiene un espectro de absorbancia media 0 y desviación estándar 1. La ecuación para el cálculo del SNV es:

$$Abs_j^{SNV} = \frac{Abs_j - \overline{Abs}}{S} \quad (4)$$

donde Abs_j es la absorbancia original a la longitud de onda j , \overline{Abs} es la absorbancia media del espectro y S es la desviación estándar.

- Derivadas: la derivación de espectros es uno de los pretratamiento más utilizados en espectroscopia NIR por su capacidad de disminuir sus problemas más característicos: solapamiento de bandas y variaciones de la línea base. La utilización de la primera derivada elimina desplazamientos constantes de la línea base en todo el espectro mientras que la segunda derivada desplazamientos que varían linealmente con la longitud de onda.

Uno de los métodos más utilizados en el cálculo de derivadas es el propuesto por Savitzky-Golay [71-72]. En la utilización de las derivadas debe tenerse en cuenta que este pretratamiento espectral disminuye la relación señal/ruido de los espectros.

2.1.1 INFORMACIÓN CUALITATIVA

Análisis en Componentes Principales (PCA)

Uno de los métodos de reducción de variables más utilizado es el Análisis en Componentes Principales (PCA, *Principal Component Analysis*) que permite, entre otros objetivos, reducir la correlación espectral y englobar las muestras analizadas en función de la variabilidad que contienen [68].

Consideremos la matriz de datos espectroscópicos \mathbf{X} , donde cada una de las I filas corresponde al espectro de una muestra, y cada columna es la absorbancia a la longitud de onda J . El objetivo del PCA es explicar la misma información original de la matriz \mathbf{X} en un número inferior de ejes o dimensiones, que corresponden a las variables de la matriz de datos. Estos nuevos ejes se llaman componentes principales. Su característica principal es que el primero recoge la máxima variabilidad espectral del conjunto de muestras, el segundo recoge una variabilidad inferior al primero y es ortogonal a él, y así sucesivamente. El número de componentes principales se ha de seleccionar de forma que contengan información relativa a las muestras y no contengan información relativa al ruido.

La matriz \mathbf{X} se descompone en el producto de dos matrices, una de scores \mathbf{T} y otra de loadings \mathbf{P} , quedando una matriz de residuales \mathbf{E} , con la información no recogida por \mathbf{T} y \mathbf{P} :

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad (5)$$

\mathbf{T} son las coordenadas de cada punto respecto a los nuevos ejes (componentes principales) y \mathbf{P} son los cosenos de los ángulos entre los nuevos ejes y los ejes originales. El conjunto de datos \mathbf{X} , descrito inicialmente por variables correlacionadas, queda ahora definido por un nuevo conjunto de variables no correlacionadas entre sí en un sistema de ejes ortogonales.

La interpretación de los resultados obtenidos con el PCA para una clasificación, 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) PC. Si existe una relación entre las muestras, en el gráfico de los *scores*, los puntos aparecerán agrupados; mientras que si las muestras no se asemejan entre si, los puntos aparecerán dispersos entre si. Esta característica permite realizar una selección de muestras que sean representativas del conjunto global de muestras y que posteriormente sean utilizadas en calibraciones. El PCA permite observar diferencias en muestras farmacéuticas en función del contenido de API, o en función de la variable que se quiera analizar, por lo que aporta información previa a la utilización de herramientas para el análisis cuantitativo.

2.1.2 INFORMACIÓN CUANTITATIVA

La información contenida en los espectros NIR puede correlacionarse con magnitudes físicas o químicas, obteniéndose ecuaciones o modelos de calibración. El fundamento de cualquier calibración es establecer una relación entre la señal instrumental y la propiedad a determinar, y posteriormente predecir esta propiedad a partir de la señal medida en una muestra desconocida. Es posible utilizar calibraciones univariantes cuando la señal es selectiva a la propiedad estudiada. Sin embargo La Espectroscopia NIR proporciona un gran número de variables respuesta para cada muestra, variables que en general no pueden ser asignadas a un solo analito. Esto ha propiciado el desarrollo de métodos de calibración capaces de relacionar múltiples variables con la propiedad a determinar. Estos métodos son conocidos como Métodos de Calibración Multivariantes. Las características principales de estos métodos son:

- Trabajan con un gran número de medidas de cada muestra, normalmente con el espectro completo.
- Son necesarios los valores de la propiedad a determinar para cada muestra.
- Es posible determinar de forma independiente cada una de las propiedades que se desea cuantificar.

Regresión Parcial por Mínimos Cuadrados (PLS)

Para el PLS, se dispone de una matriz de datos espectrales \mathbf{X} y una matriz \mathbf{Y} que contiene los datos de la propiedad a determinar. La regresión PLS se basa en un cálculo de reducción de variables, donde las matrices \mathbf{X} (espectral) e \mathbf{Y} (analito) se descomponen simultáneamente según:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (6)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (7)$$

Donde las matrices \mathbf{T} y \mathbf{U} son, respectivamente, las matrices de scores de los bloques \mathbf{X} e \mathbf{Y} . Las matrices \mathbf{P}^T y \mathbf{Q}^T son las matrices de loadings de los bloques \mathbf{X} e \mathbf{Y} , y \mathbf{E} y \mathbf{F} son las matrices de residuales del bloque \mathbf{X} y del bloque \mathbf{Y} respectivamente.

En PLS, al igual que en PCA, se han de seleccionar el número de factores adecuado al sistema, recogiendo información cuantitativa pero sin recoger ruido o información espectral relativa a otros parámetros no relacionados con la calibración.

La descomposición de las dos matrices no se hace por separado, sino de forma simultánea. La principal característica de esta descomposición es que busca la máxima correlación entre los espectros y la propiedad a determinar. Esta descomposición implica pérdida de ortogonalidad entre los factores pero mejora la capacidad predictiva del modelo de calibración.

Una vez se ha establecido el modelo de calibración correcto, es posible predecir el resultado para una nueva muestra o para un conjunto de muestras externo a la calibración. La correcta predicción de las nuevas muestras dependerá de la buena capacidad predictiva del modelo de calibración.

Classical Least Squares (CLS)

La regresión Clásica de mínimos cuadrados es un método de cuantificación ampliamente conocido que supone que cada medición es la suma ponderada de señales linealmente independientes [73].

CLS es un método directo que requiere un total conocimiento de todos los componentes de las muestras, en el caso de la Espectroscopia NIR, los espectros de todos los componentes, de las muestras de calibración, y funciona correctamente en sistemas muy sencillos.[74].

El modelo CLS asume que el espectro de cada muestra es la suma de los espectros de puro componente ponderado por su concentración. En esencia, la matriz \mathbf{X} se descompone como sigue:

$$\mathbf{X} = \mathbf{P}\mathbf{S}^T \quad (8)$$

donde \mathbf{P} es la concentración de la matriz por cada componente puro y \mathbf{S} corresponde a los espectros. El verdadero interés en CLS se basa en calcular por método directo la concentración de nuevas muestras a partir de los espectros puros de los componentes. De esta forma, \mathbf{P} se puede calcular fácilmente como sigue:

$$\mathbf{P} = \mathbf{X}\mathbf{S}^+ \quad (9)$$

donde \mathbf{S}^+ corresponde a la pseudo-inversa de la matriz \mathbf{S} :

$$\mathbf{S}^+ = \mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (10)$$

A pesar de la facilidad de cálculo de esta concentración de la matriz, el CLS tiene dos principales inconvenientes: la señal puede verse afectada por variaciones no sistemáticas que influyen en el patrón espectral, y que no siempre se dispone de los espectros puros de todos los componentes. En el caso de muestras naturales (por ejemplo agroalimentarias) no puede utilizarse este método, mientras que para el caso que se muestra en esta memoria, donde se utilizan muestras farmacéuticas, se ha podido utilizar al disponer de los espectros puros.

Resolución multivariante de curvas MCR-ALS

La resolución multivariante de curvas es una de las técnicas quimiométricas de más reciente aplicación. Su desarrollo teórico se ha producido a lo largo de la década de los 70, y no ha sido hasta finales de los 90, cuando se ha aplicado en diferentes campos de

la química analítica [75,76]. Hasta finales de los 90, MCR-ALS estaba considerada como una técnica cualitativa donde el interés se centraba en obtener información espectral de los compuestos de las muestras estudiadas. Actualmente, MCR-ALS empieza a considerarse como una técnica con capacidad para obtener información cuantitativa del sistema estudiado, siendo interesante respecto a otras técnicas quimiométricas, anteriormente comentadas, porque es capaz de resolver cualitativa y cuantitativamente sistemas químicos con un mínimo de información externa.

La resolución multivariante de curvas asume que los datos experimentales siguen un modelo lineal que de forma matricial, se representa como [77]:

$$D = CS + E \quad (11)$$

Donde **D** corresponde a la matriz espectral, **C** es la matriz calculada correspondiente a las concentraciones de los analitos involucrados en la matriz de espectros inicial, **S** corresponde a la matriz de espectros recalculada de los analitos y finalmente **E** corresponde a una matriz de residuales y representa toda la información no explicada por las matrices **C** y **S** y que está presente en la matriz **D**.

Gráficamente, la ecuación 11 se puede representar como:

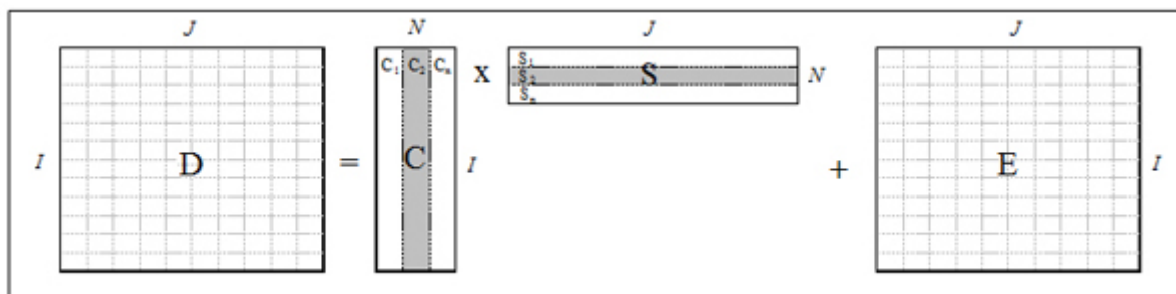


Figura 2.2. Representación gráfica de MCR.

Donde I es el número de espectros registrados o número de muestras, J es el número de variables espectrales y N es el número de especies químicas involucradas en la reacción.

Cuando el modelo lineal se cumple, d_{ij} es la suma de la absorción de las N especies presentes en la muestra I a la longitud de onda J

$$d_{ij} = \sum_n c_{in} s_{nj} + e_{ij} \quad (12)$$

Una vez conocido el número de analitos involucrados en la matriz de datos, el objetivo de los métodos multivariantes de resolución de curvas es resolver el sistema representado en la ecuación 11 utilizando solo la matriz \mathbf{D} , es decir, obtener información cualitativa y cuantitativa solo a partir de datos espectroscópicos.

La matriz de datos \mathbf{D} se descompone mediante la metodología de Mínimos Cuadrados Alternados (ALS). La solución de la descomposición no es única ya que se pueden obtener una gran número de combinaciones de \mathbf{C} y \mathbf{S} , que multiplicadas entre sí permitan obtener la misma matriz \mathbf{D} .

Estas ambigüedades pueden ser corregidas mediante la aplicación restricciones en el cálculo, para evitar soluciones ilógicas experimentalmente aunque lógicas matemáticamente, como pueden ser concentraciones negativas o espectros negativos (lo cual dependerá del sistema de registro o de los pretratamientos utilizados). También existen problemas de ambigüedad en las soluciones de la ecuación 11, que aumentan aún más el número de soluciones al sistema:

- **Ambigüedad rotacional:** ocurre cuando no existe selectividad y cuando hay dos o más componentes con una elevada correlación. El espectro calculado para cualquiera de ellos será una combinación lineal desconocida de los verdaderos espectros. Teniendo en cuenta que la absorción en el infrarrojo próximo está formada por bandas anchas y solapadas, donde prácticamente no existen

regiones de absorción selectivas, la ambigüedad rotacional es uno de los problemas más importantes en la resolución de estos sistemas.

- **Ambigüedad de intensidad:** Esto significa que los espectros y los perfiles de concentración calculados pueden estar multiplicados por cualquier número. No es un problema serio en análisis cualitativo pero sí en análisis cuantitativo. En ambos casos de ambigüedades es necesaria aportar información adicional a la matriz **D** para resolver esta ambigüedad.

Restricciones aplicadas durante el proceso de iteración

- **Restricción de no negatividad:** las concentraciones de las especies químicas siempre han de ser positivas o cero. Los espectros generalmente son positivos, excepto si se han utilizado pretratamientos espectrales como las derivadas o el SNV. Esta restricción se suele aplicar forzando a los valores negativos a ser cero para continuar con el proceso de iteración.
- **Restricción de unimodalidad:** esta restricción es aplicable a perfiles espectrales o de concentración que solo posean un máximo. En primer lugar se encuentra el valor máximo, y los valores a un lado y otro del máximo han de disminuir de forma monótona.
- **Closure:** aplicable en sistemas químicos cerrados de forma que la suma de concentraciones de todas las especies químicas en cada punto de la reacción es constante.
- **Convergencia:** el proceso de iteración se para cuando se cumple el criterio de convergencia establecido o cuando se alcanza un número de iteraciones máximo.

Versión Aumentada MCR-ALS

Cuando una matriz espectral no puede ser correctamente descompuesta debido a deficiencias espectrales, una de las opciones puede ser aportar información

directamente a la matriz espectral [78]. Si se disponen las matrices una encima de la otra, con las columnas en común, se obtiene una matriz **D** aumentada:

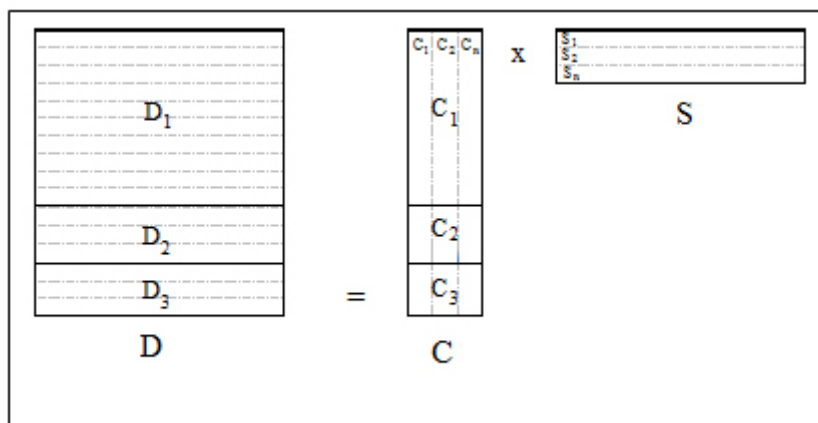


Figura 2.3. Representación gráfica de la versión aumentada de MCR.

La matriz **C** calculada corresponde a los perfiles de concentración de las especies presentes en las submatrices de **D**, es decir, el sistema no fuerza a que los perfiles de concentración sean iguales, una misma especie puede tener diferentes perfiles espectrales en cada matriz C_i .

No ocurre lo mismo para la matriz **S**. Los espectros calculados para cada especie han de ser iguales para las diferentes matrices. El tratamiento conjunto de matrices aporta ventajas en el cálculo de MCR-ALS:

- El número de soluciones de la ecuación [8] disminuye.
- Se puede resolver la ambigüedad de intensidad utilizando varias matrices con diferentes condiciones iniciales de concentración en un cálculo conjunto.

2.2 HERRAMIENTAS PARA EL ANÁLISIS DE IMAGEN

Las técnicas de clasificación utilizando clustering han sido utilizadas únicamente en el estudio realizado con la técnica de Imagen NIR, por lo que la teoría a continuación está descrita exclusivamente para su aplicación a datos hiperespectrales.

La identificación y clasificación de productos farmacéuticos implicados es uno de los ensayos rutinarios necesarios para el control del proceso farmacéutico. Con la finalidad de encontrar mapas que puedan dar una visión de la distribución de los analitos que forman las muestras se han aplicado las herramientas de clustering, además de utilizar el coeficiente de correlación, como herramienta innovadora en el tratamiento de imágenes al poder obtener tantos coeficientes de correlación como píxeles que contiene una imagen de manera sencilla para ser representados en formato de mapa de correlaciones.

CLUSTER ANALYSIS

El análisis de clusters se ajusta a una serie de métodos de clasificación que realizan el análisis de la imagen por la segmentación de los datos específicos de la composición química basada en la información de los píxeles. La característica principal de estos métodos consiste en agrupar los objetos (píxeles cuando hablamos de imágenes) en clases [79].

Antes de empezar el análisis, puede ser recomendado para reducir el número de variables a un mínimo suficiente mientras que la máxima cantidad de información que se mantiene (extracción de características). Esto permitirá mejorar la eficiencia computacional y la reducción del número de clases posibles. Entre la propuesta de métodos de extracción de características [80], una buena alternativa es utilizar los resultados obtenidos en un anterior análisis de PCA como entrada [79].

Número de clusters

El número de clusters necesarios para clasificar las imágenes es uno de los aspectos más a tener en cuenta antes de aplicar cualquier técnica de clasificación. En esta memoria se presentan dos ecuaciones para el cálculo del número de clusters.

Silhouette index

Este parámetro [81] se calcula para cada x_{mn} píxel proporcionando una medida acerca de la similaridad entre los puntos de un mismo cluster comparados con los otros clusters. (Eq. 14):

$$S_{mn} = \frac{b_{mn} - a_{mn}}{\max(a_{mn}, b_{mn})} \quad (14)$$

donde a_{mn} es la distancia media entre cada mn th píxel y los píxeles incluidos en el mismo cluster, y b_{mn} , la distancia media mínima entre cada mn th píxel y los píxeles incluidos en otros clusters. Valores de S_{mn} cercanos a 1 confirman que los píxeles están correctamente asociados a los clusters, mientras que valores cercanos a -1 corresponde a una incorrecta clasificación.

Entre los diferentes algoritmos de clasificación de análisis de imagen [82], nos hemos centrado en los algoritmos *K-means cluster analysis* y el *Fuzzy C means cluster analysis*.

Algoritmo K-means (KM)

El algoritmo K-means (KM) asigna cada píxel de la imagen al cluster cuyo centro está más cercano al minimizar la suma de los cuadrados de las distancias de cada uno de los píxeles a su correspondiente centro (Eq. 12):

$$J = \sum_{k=1}^K \sum_{mn=1}^{MN} \|x_{mn}^{(k)} - m_k\|^2 \quad (15)$$

Donde $x_{mn}^{(k)}$ es el píxel y m_k es el centro del cluster k . Los pasos de este algoritmo son los siguientes: [83]:

- 1) Escoger el número de clases-clusters.
- 2) Generar k clusters y determinar los correspondientes centros
- 3) Asignar cada píxel al centro más cercano.
- 4) Determinar J y recalcular los nuevos centros.

5) Repetir los puntos 3 y 4 hasta llegar al criterio de convergencia.

La principal ventaja de este método es la simplicidad, que permite trabajar con un gran número de datos. Sin embargo, esta clasificación asigna únicamente un cluster a un píxel.

Fuzzy C-means algorithm (FCM)

A diferencia del anterior algoritmo FCM asigna a cada píxel de la imagen un grado de pertenencia a todos los clusters simultáneamente. En este sentido, esta clasificación permite observar diferencias entre píxeles del centro de un cluster con los píxeles que se encuentran en el límite entre dos clusters [79]. El grado de pertenencia de un píxel a cada cluster se calcula:

$$u_{mnk}^g = \left[\sum_{i=1}^C \left(\frac{\|x_{mn}^{(k)} - m_k\|^2}{\|x_{mn}^{(k)} - m_i\|^2} \right)^{1/(g-1)} \right]^{-1} \quad (16)$$

donde g corresponde a la constante “*fuzzifier*” que determina el grado de pertenencia y varía de 0 a 1.

Ahora la ecuación 5 se amplía:

$$J = \sum_{k=1}^K \sum_{mn=1}^{MN} u_{mnk}^g \|x_{mn}^{(k)} - m_k\|^2 \quad (17)$$

Y el centro de cada cluster se calcula:

$$m_k = \frac{\sum_{mn=1}^{MN} u_{mnk}^g x_{mn}}{\sum_{mn=1}^{MN} u_{mnk}^g} \quad (18)$$

Los pasos a seguir para este algoritmo son muy parecidos a los descritos en el apartado anterior con la ventaja de minimizar la varianza entre clusters debido a la aportación del grado de pertenencia.

MAPAS DE CORRELACIÓN

Dentro de las técnicas de imagen química, los coeficientes de correlación aportan información sobre la composición química de cada píxel y los mapas de correlación aportan información visual acerca de la distribución de cada analito en la mezcla. El aumento de la proporción de un analito en una muestra supondrá una mayor contribución espectral, y por tanto, un valor superior de coeficiente de correlación

Desde nuestro conocimiento sólo en una ocasión se ha descrito la aplicación del coeficiente de correlación en la determinación de un analito a partir de imágenes hiperespectrales NIR [84] y en otro trabajo [85] la determinación de un API en diferentes muestras procedentes de una proceso de producción a partir de espectros NIR convencionales.

Al trabajar con imágenes hiperespectrales se calcula los coeficientes de correlación entre los espectros puros y los espectros de cada píxel de la imagen obteniendo un mapa de coeficientes de correlación para cada analito. El coeficiente de correlación se calcula:

$$corrcoef = \frac{\sum_{\lambda} x_{mn\lambda} y_{\lambda}}{\sqrt{\sum_{\lambda} x_{mn\lambda}^2 \sum_{\lambda} y_{\lambda}^2}} \quad (13)$$

donde, $x_{m\lambda}$ es el espectro de cada m th píxel e y_λ corresponde al espectro puro de cada analito. El uso de coeficientes de correlación en la imagen química puede aportar información acerca de la distribución de los analitos en la superficie de la muestra.

2.3 OTROS

Bibliotecas Espectrales mediante el Coeficiente de Correlación

Finalmente, la última herramienta utilizada en esta Tesis no ha sido clasificada en ninguno de los apartados anteriores.

Una biblioteca NIR de identificación debería englobar todas las materias primas utilizadas por el fabricante con el fin de ser capaz de identificar todas las posibles sustancias y evitar o reducir al mínimo los errores. El método que se utilizará debe permitir la identificación inequívoca de cada compuesto en la actualidad la biblioteca y la exclusión de los no presentes. También debe ser capaz de distinguir muy similares entre los compuestos utilizados en diferentes aplicaciones (por ejemplo productos con diferentes tamaños de partículas, polimorfos, los productos en diferentes grados y de diferentes proveedores).

El coeficiente de correlación ha sido utilizado como parámetro cualitativo aportando el grado de similitud del espectro de una muestra con los espectros puros de los analitos que componen la muestra [86]. La identificación de compuestos mediante la utilización de una biblioteca espectral que contiene los espectros de un elevado número de compuestos se puede realizar mediante la aplicación del coeficiente de correlación (CC) como criterio discriminante.

Cuando el CC no es un parámetro suficientemente discriminante entre dos analitos, existen ambigüedades que deben ser resueltas. En esta tesis, se ha utilizado la Distancia de Mahalanobis basados en cálculos sobre el espacio de los componentes principales. De un modo análogo, cada clase se define como una elipse obtenida a partir de un PCA.

La identificación de la muestra se realiza a través de la diferencia entre el centroide de la clase con la muestra desconocida [87].

2.4 ADAPTACIÓN A TÉCNICAS DE IMAGEN

Una de las principales características de las imágenes hiperespectrales es la gran cantidad de información recogida de una muestra. En NIR Imagen, por ejemplo, un espectro NIR se puede medir en un amplio rango de longitud de onda en cada píxel de la superficie. Esta estructura puede entenderse como una matriz tridimensional X ($M \times N \times \lambda$) (figura 2.4), donde M y N representan los ejes espaciales y el eje λ corresponde al patrón espectral. Atendiendo a la estructura de la matriz obtenida X , puede ser definido como cubo tridimensional [83], pero también es común para referirse a esta matriz como una imagen hiperespectral [88]. Esas estructuras de datos pueden estar compuestas por miles de espectros. Por lo que hace evidente que esta cantidad de información espectral, normalmente muy correlacionados entre sí, necesite de los tratamientos matemáticos de gran alcance para extraer la información conveniente de los datos primarios.

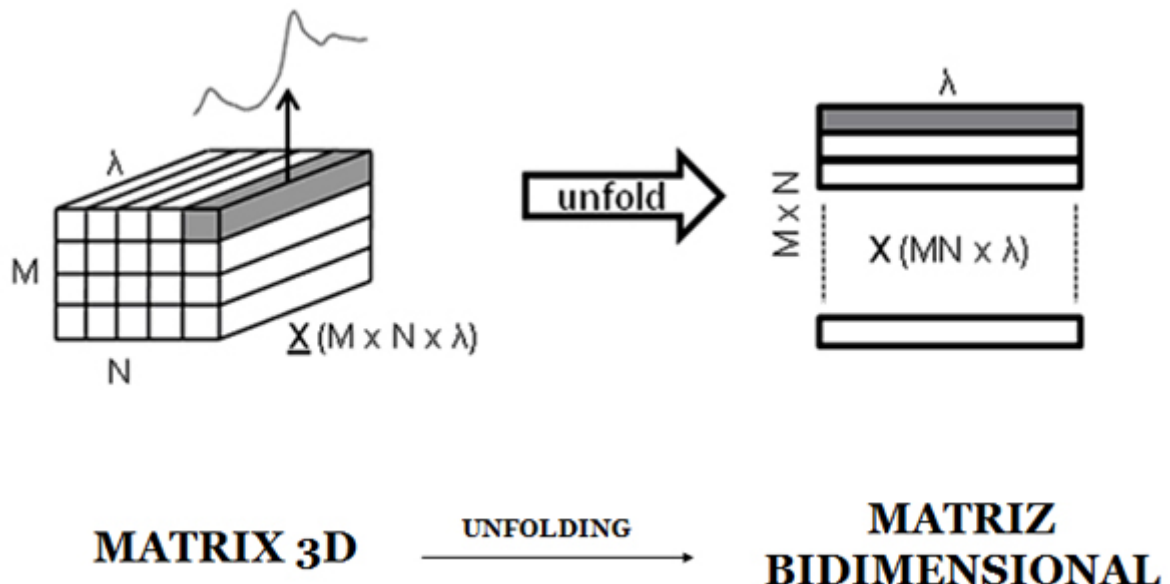


Figura 2.4 Representación del desdoblamiento de la matriz tridimensional

Unfolding

Las herramientas quimiométricas descritas en el apartado anterior no pueden ser aplicadas directamente a matrices tridimensionales, por lo que el desdoblamiento o *unfolding* es un paso necesario para adaptar las matrices tridimensionales a matrices de 2 dimensiones.

Como ya se ha comentado en anteriores apartados, cuando se registran imágenes se obtiene una matriz tridimensional \mathbf{X} ($M \times N \times \lambda$). Cuando se realiza el desdoblamiento (*unfolding*) se obtiene una matrix \mathbf{X} con dimensiones ($MN \times \lambda$). Esta etapa es crucial para la posterior aplicación de algoritmos bidimensionales como PCA, clustering, CLS y MCR.

3. OBJETIVO de la Tesis

La producción farmacéutica tiene por objetivo proporcionar un fármaco que cumpla unas estrictas condiciones de calidad que aseguren la satisfacción del usuario; este es un proceso complejo y variado que está constituido por diferentes etapas. La garantía de calidad del producto final se puede conseguir mediante un continuado control en cada una de las diferentes etapas, que obligan a la aplicación de muy diferentes técnicas instrumentales capaces de determinar no sólo la composición química de los preparados intermedios sino también de características físicas de los mismos. La realización de estas determinaciones consume un tiempo considerable, prolongando el proceso y reduciendo la productividad de la empresa. Todo ello ha conducido a la Food and Drug Administration (FDA) a promover una iniciativa, Process Analytical Technology (PAT), que tiene por objeto incorporar al proceso productivo nuevas tecnologías que aporten un mejor y más amplio conocimiento del proceso productivo, reduzcan el tiempo de control y aumenten la productividad.

Se han propuesto una amplia variedad de técnicas instrumentales capaces de aportar soluciones a este reto, pero sin duda es la Espectroscopía NIR la más adecuada por sus características propias tales como: rapidez, no preparación de muestra, determinación de composición y de propiedades físicas, facilidad de incorporación al proceso productivo, etc.

Con el objeto de mejorar diferentes etapas relacionadas con los procesos de producción farmacéuticos que permitan asegurar la calidad final del fármaco y también de etapas intermedias del proceso, hemos abordado el estudio de la aplicación del NIR en algunas situaciones de interés del proceso de fabricación mediante la aplicación de sistemas de espectroscopía NIR convencionales y de técnicas de Imagen Química de más reciente implantación.

El objetivo de esta tesis es desarrollar nuevas aplicaciones de la espectroscopía NIR que puedan tener interés en el proceso productivo de fármacos con el objeto de ampliar el conocimiento del mismo y que puedan mejorar la calidad tanto del producto final como

del proceso de fabricación. La finalidad última es que las aplicaciones desarrolladas se implanten en el proceso productivo, bien en el control de fases intermedias bien en el del producto acabado, que serán la demostración palpable de que se han cumplido los objetivos propuestos.

3. OBJECTIVES

Pharmaceutical production aims to provide a drug that meets strict quality conditions that ensure user satisfaction. This is a complex and varied procedure constituted by different phases. Quality assurance of final product can be achieved through a continuous control in the different stages, which require the application of different instrumental techniques capable of determining not only the chemical composition of the preparations but also for intermediate physical characteristics of them. These determinations consume considerable time reducing the productivity. This has led to the Food and Drug Administration (FDA) to promote an initiative, Process Analytical Technology (PAT), which aims to incorporate new technologies to the production process to provide a better and wider knowledge of the production process, reducing the time control and increase productivity.

A wide variety of instrumental techniques capable of providing solutions to this challenge have been proposed, but certainly the NIR spectroscopy is the most appropriate technique for their characteristics such as speed, no sample preparation, determination of composition and physical properties, ease incorporation of the productive process, and so on.

In order to improve various stages associated with pharmaceutical production processes to ensure the final quality of the drug and also intermediate stages of the process, we have approached the study of the application of NIR in some situations of interest to the manufacturing process through implementing systems to conventional NIR spectroscopy and imaging techniques Chemical latest deployment.

The objective of this thesis is to develop new applications of NIR spectroscopy that may be interesting in the production process in order to broaden awareness of it and that can improve the quality of the final product as the manufacturing process. The ultimate aim is that the applications developed are located in the production process, either in

OBJECTIVES

control or in the intermediate stages of the finished product, which will be the demonstration that it has met its objectives.

3. METODOLOGÍA Y DISCUSIÓN DE LOS RESULTADOS

4. METODOLOGÍA Y DISCUSIÓN DE LOS RESULTADOS

A continuación se describen las características más relevantes de la metodología empleada para el desarrollo de los nuevos métodos propuestos, subdividido en 3 apartados: metodología experimental donde se describen las muestras analizadas y los métodos de preparación de muestras, la metodología instrumental empleada y finalmente el tratamiento de los datos desde la obtención del espectro NIR hasta la obtención del resultado final.

4.1 METODOLOGÍA EXPERIMENTAL

A lo largo de esta memoria se han analizado muestras de productos farmacéuticos que presentan distintas características físicas: muestras en polvo, granuladas, comprimidos y geles.

La determinación de parámetros de interés a partir del espectro NIR de las muestras se ha realizado mediante métodos de calibración multivariante. La metodología a seguir comprende diferentes etapas: recopilar un conjunto de muestras representativas que incluya todas las posibles fuentes de variabilidad del proceso de fabricación y que abarque un intervalo de valores del parámetro objetivo de estudio lo suficientemente amplio. El espectro NIR depende de la composición química de la muestra, pero también depende de algunas propiedades físicas tales como el tamaño, la forma, la distribución de partículas y el grado de compresión de la muestra. Estas diferencias físicas pueden dar lugar a efectos de scattering complican los modelos de calibración en detrimento de la calidad de los resultados de los análisis. Estos efectos pueden reducirse mediante la aplicación de pre-tratamientos espectrales adecuados y modelados para asegurar que el conjunto de calibración incluye todas las posibles fuentes de variabilidad en las muestras.

Preparación de muestras

Los procesos de producción de fármacos siguen unos estrictos protocolos durante su fabricación por lo que la variabilidad que puede haber entre varios lotes de producción es mínima. Como valor máximo la gran mayoría de farmacopeas permiten obtener el producto final con una variación de $\pm 5\%$ de todo el valor nominal de API.

Sin embargo, uno de los requisitos para realizar calibraciones correctas y robustas consiste en obtener un conjunto de muestras de calibración que contengan un amplio intervalo de la propiedad que se quiere analizar.

Con el objetivo de obtener conjuntos de calibración válidos para poder analizar correctamente las propiedades deseadas en el producto farmacéutico, en la presente Tesis se han propuesto diferentes estrategias para ampliar la variabilidad en las muestras que se utilizaran en la construcción de los modelos de calibración y que permitirán crear modelos robustos:

Sobre e Infra dosificar muestras de la producción

Esta metodología se aplica habitualmente a muestras en polvo y consiste en añadir, a muestras de producción pequeñas cantidades de API si se quieren obtener muestras con una mayor concentración de API y de mezcla de excipientes cuando se quieren preparar muestras con un contenido inferior de API. Las adiciones de API o excipientes suelen ser reducidas, por lo que los cambios en las propiedades físicas de las muestras que producen estas adiciones no representan cambios significativos en la matriz de las muestras originales. Este es un medio eficaz y simple y permite obtener la concentración adecuada de API para cubrir el rango de concentraciones deseado. Además, las propiedades físicas de las nuevas muestras pueden suponerse esencialmente idénticas a la de las muestras de la producción.

Muestras sintéticas de laboratorio:

Se preparan por pesada de cantidades conocidas de API y de los excipientes en polvo, y mezclados hasta llegar a la homogeneidad. Este procedimiento tiene algunas ventajas. Es menos laborioso y costoso que el anterior método, pudiendo obtener fácilmente muestras con la concentración exacta de API y excipientes requeridos. El método no precisa de la utilización de un método de referencia para determinar el contenido de API y la concentración se conoce con mayor exactitud (es un método absoluto). Sin embargo, las muestras de laboratorio pueden diferir de las muestras de la producción real en algunas características físicas como consecuencia de las operaciones realizadas durante el proceso de producción. Este método permite obtener modelos simples y apropiados para estas muestras pero habitualmente conduce a errores en la predicción de muestras producción. Esta deficiencia, sin embargo, puede ser solucionada mediante la inclusión de un número suficiente de muestras de producción en el conjunto de calibración. Este procedimiento ha demostrado su eficacia en la labor realizada por los autores grupo de investigación para la determinación del contenido de API en algunos productos farmacéuticos. En algunos casos, sin embargo, las diferencias espectrales entre las muestras de laboratorio de fabricación y producción son tan grandes que los modelos resultantes son bastante complejos y poco robustos.

Para resolver el problema de las diferencias físicas entre las muestras preparadas en el laboratorio y las muestras de producción, en la presente Tesis se ha estudiado la aplicación de una simulación del proceso de producción a un conjunto de muestras preparadas en el laboratorio para obtener finalmente unas muestras de laboratorio semejantes en sus características físicas a las muestras de producción.

La preparación de muestras granuladas en el laboratorio se realizó tomando alícuotas de las muestras preparadas en el laboratorio de unos 5 gramos a las que se añadieron entre 1,5 – 3 g de agua mediante un pulverizador. La pasta húmeda obtenida se mezcló manualmente con una espátula en un mortero de vidrio durante 20 minutos hasta

obtener una masa granulada; finalmente se introdujo la masa en una estufa de vacío a unos 60 °C durante 5 horas (VACIOTEM, P-Selecta, Abrera, Spain) para su secado. Una vez seco este granulado se molturó manualmente hasta obtener un tamaño de partícula similar al del producto comercial.

A continuación se detallan las características de las muestras de producción utilizadas en la presente tesis:

- Enantyum sobres

La preparación farmacéutica es un granulado sólido, que contiene como API Dexketoprofeno Trometamol (DKP) (10 mg·g⁻¹), Sacarosa (Suc) como excipiente mayoritario (970 mg·g⁻¹) y otros excipientes minoritarios: Aroma de Limón (Fla), Neohesperidina dihidrochalcone (Neo), Amarillo de Quinoleína (Qui) y Glicirrinato amónico (Gly). Para analizar estas muestras fue necesario preparar muestras en polvo en el laboratorio que posteriormente fueron granuladas.

- Enantyum Comprimidos

La formulación farmacéutica se presenta un contenido de 135 mg/g de Dexketoprofeno Trometamol, Celulosa Microcristalina (500 mg/g) como excipiente mayoritario y otros, en menor proporción, que incluyen Almidón de maíz, Carboximetilalmidón y Diestearato de Glicerol. La mezcla de API, celulosa microcristalina y almidón de maíz es sometida a un proceso de granulación y posterior adición de los otros excipientes. Finalmente el producto es compactado para formar comprimidos. Finalmente el comprimido se recubre de una fina película de laca (compuesta por Laca OPADRY (85 %) and Propilenglicol (15%) diluido en Agua Desionizada. En este estudio se prepararon muestras por sobre e infra dosificación.

- Fastum

Las muestras de este preparado farmacéutico están constituidas por un gel hidroalcohólico que contiene como principio activo Ketoprofeno (KET, 25 mg·g⁻¹), dos conservantes p-hidroxibenzoato de metilo (MET, 0.8 mg·g⁻¹) y de propilo (PRO, 0.2 mg·g⁻¹). El agua y el etanol representan los excipientes mayoritarios (> 950 mg·g⁻¹). En este estudio se prepararon muestras de laboratorio.

Proceso de granulación monitorizado

El proceso de granulación monitorizado en la presente Tesis se ha llevado a cabo en un reactor de lecho fluido (Fluid Bed System GLATT WSG300). El proceso habitual de granulación para este producto farmacéutico se realiza con una carga de 260 Kg de materia prima, API y los dos excipientes mayoritarios (celulosa microcristalina y almidón), en las condiciones de trabajo actualmente definidas para la obtención del granulado de las características deseadas. La carga de la materia prima se realiza por aspiración en la zona superior del reactor manteniendo un flujo de aire que mantiene la carga en agitación constante con el objeto de tener una mezcla homogénea de todos los componentes antes de iniciar propiamente el proceso.

Un 27 % del total de Almidón de Maíz es utilizado para preparar con 200 L de agua destilada la solución ligante que se añadirá durante la etapa de granulación.

Finalizada la carga de los reactivos empieza el proceso, que tiene una duración de 255 minutos, y que consta de tres etapas: agitación de la carga por un flujo de aire caliente (15 min.), adición de la solución ligante (140 min.) y finalmente la etapa de secado (100 min.) En la primera etapa, la temperatura del flujo de aire de agitación es de 60 °C, y tiene como objetivo completar la homogeneización de toda la muestra antes de la granulación; a continuación, fijando una temperatura del aire de 80°C se empieza la adición de solución ligante (suspensión de almidón de maíz en agua) con un flujo de 1,5 L/min por aspersión en la zona superior del reactor para evitar que se formen gránulos de gran tamaño que obturen la salida de aire y a la vez que se incorpore

lentamente la solución ligante con toda la muestra. La última etapa del proceso es el secado, que se realiza con una temperatura del aire ascendente de 90 °C, y que tiene como finalidad evaporar el agua añadida en la etapa anterior. El flujo de aire se mantiene hasta el secado completo de la masa granulada.

En el estudio de NIR Imagen se prepararon 4 muestras binarias (Ibuprofeno y Almidón) en el laboratorio como se observa en la figura X:



Figura 4.1. Representación gráfica de las 4 muestras binarias donde se aprecian las regiones de cada analito puro y la región de mezcla con la relación de los 2 analitos.

3 muestras presentan 3 regiones perfectamente definidas: 2 para los analitos puros, y una región de mezcla donde varía la concentración en función de la muestras; la última muestra corresponde a una mezcla 65:35 Ibuprofeno y Almidón agitada correctamente.

4.2 METODOLOGÍA INSTRUMENTAL.

A continuación se detallan la instrumentación empleada para el desarrollo de esta memoria. Además de diferentes equipos de espectroscopia NIR, ha sido necesaria la

utilización de otros instrumentos aplicados en la realización del método de referencia para construir modelos de calibración.

4.2.1 INSTRUMENTACIÓN NIR

NIR FOSS 5000

El espectrofotómetro tiene incorporado un módulo RCA (Rapid Content Analyzer) que permite registrar espectros en modo reflectancia y transfectancia.

Este instrumento, que registra espectros por barrido, utiliza como fuente de luz una lámpara halógena de filamento de tungsteno, que proporciona un espectro continuo entre 320 y 2500 nm. La lámpara posee un regulador de voltaje para asegurar la estabilidad de la radiación emitida. El monocromador está constituido por una red de difracción cóncava que al girar permite enfocar cada longitud de onda en la rendija del instrumento; el sistema permite un registro rápido del espectro (menos de 1 seg) con un intervalo espectral de 2 nm. El detector es de PbS y permite el registro de espectros en el rango de 1100-2500nm.

Previamente al registro de las muestras se registra un blanco constituido por una placa de cerámica. Cada espectro es el promedio de 32 scans.

Todas las muestras registradas con el NIR FOSS se realizan por triplicado, y promediando para obtener un único espectro.

Las muestras en polvo o granuladas se han registrado en modo reflectancia una cubeta de cuarzo como soporte de la muestra. El espectro de comprimidos fue obtenido mediante el registro directo sobre la ventana del instrumento en modo Reflectancia.

Los espectros NIR de muestras de muestras en gel se han registrado en modo Transfectancia utilizando la misma cubeta de cuarzo y con un reflector de oro que fija el camino óptico en 1 mm. La referencia se registró utilizando la cubeta vacía y el mismo reflector de oro.

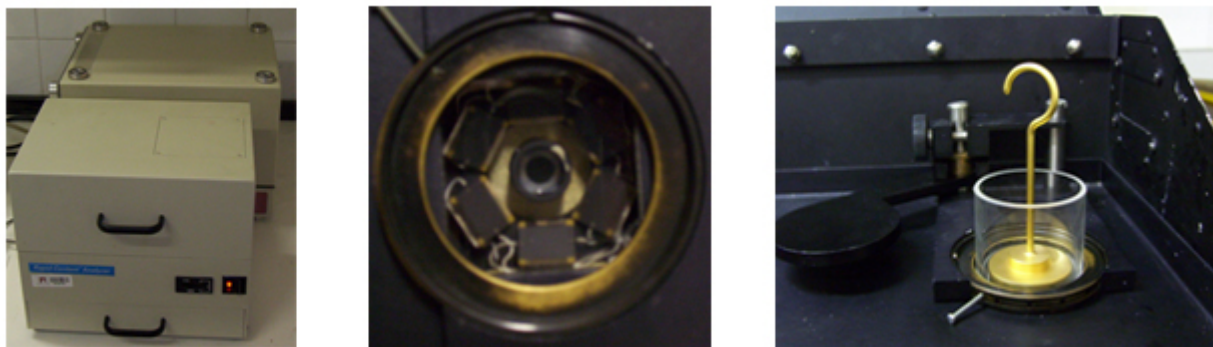


Figura 4.2. a) Imagen del Espectrofotómetro. b) Ventana del equipo. 6 detectores y haz de luz de radiación NIR. c) Montaje del sistema de medida por transfectancia mediante el uso de un reflector de oro.

NIR Spectrometer QualitySpec ASD

Este espectrofotómetro permite el registro de medidas en reflectancia y transmitancia con la ventaja de ser un instrumento portátil y por tanto puede ser trasladado a diferentes puntos de un proceso de producción para el registro *at* y *on line*. El rango espectral de trabajo del espectrofotómetro es de 350-2500 nm, aunque para la presente tesis, se utilizó exclusivamente el intervalo 1000-2500 nm que corresponde a la región del Infrarrojo cercano. El equipo consta de 3 detectores en función de los intervalos de longitudes de onda. Además, una fuente de luz y una sonda de fibra óptica hacen de éste, un equipo fácilmente portátil. Cada espectro ha sido obtenido del promedio de 100 scans registrados en un intervalo de tiempo de 10 segundos.

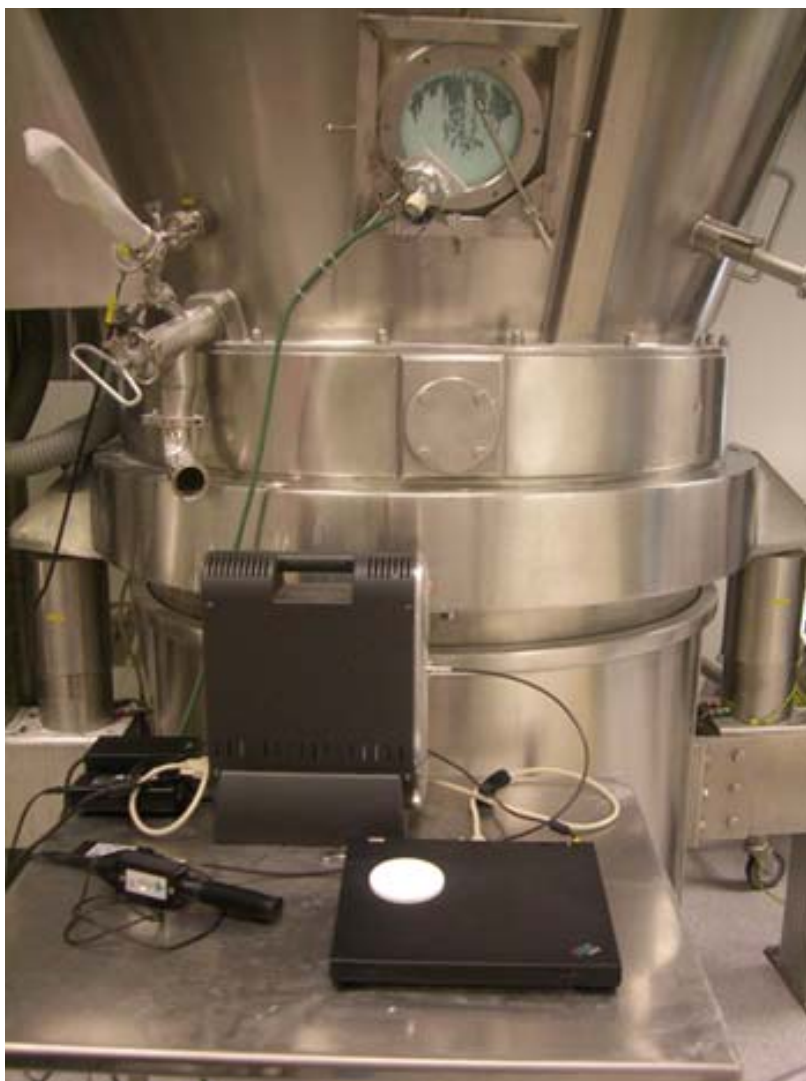


Figura 4.3 Montaje del sistema de registro durante el proceso de granulación. La sonda de fibra óptica se coloca, junto con la fuente de luz, en la ventana del reactor.

Este equipo ha sido utilizado en el estudio de monitorización de un proceso de granulación. La figura 4.3 muestra una fotografía del montaje empleado en el granulador de lecho fluido junto con el equipo de registro. Este está formado por un espectrofotómetro conectado a una computadora que almacena todos los espectros registrados. Todas las lecturas de espectros se realizaron a través de una ventana del granulador, de forma no invasiva, colocando la sonda de fibra óptica directamente enfocada a la ventana. La fuente de radiación de luz es una lámpara halógena orientada 45° respecto a la sonda.

Para el registro individual de muestras sólidas, la sonda de fibra óptica se inserta en un adaptador del propio instrumento aportando la luz necesaria para el análisis. En este tipo de registros, el dispositivo de la sonda y luz consta de una ventana que se pone en contacto con la muestra antes del registro del espectro. Un espectro de referencia debe ser obtenido antes de cualquier registro de muestra, mediante una placa cerámica.

Cámara NIR Hyperespectral



Figura 4.4 Cámara hyperespectral RODA NIR 25

Todas las imágenes fueron registradas en un espectrofotómetro de Imagen NIR “Think Spectrally Roda-25” (Think Spectrally, Valencia, Spain). Este instrumento es un *Focal Plane Array* (FPA) provisto de un detector de Mercurio Cadmio Telurio (MCT) que permite registrar imágenes en un rango espectral de 1000-2400 nm con resolución espectral de 7 nm. En las condiciones de trabajo, el tamaño de cada píxel es de

aproximadamente de 100 x 100 μm , obteniendo un área de imagen aproximada de 32.0 x 25.6 mm.

Este instrumento mide en modo de reflectancia difusa de espectro continuo con un tiempo total aproximado de 2 minutos por imagen. La iluminación de la muestra es un parámetro relevante en la obtención de imágenes de calidad. Por ello, 4 lámparas halógenas están colocadas enfocando a la muestra con una orientación de unos 45 °.

4.2.2 TÉCNICAS INSTRUMENTALES DE REFERENCIA

En los trabajos que se presentan en esta tesis se han utilizado varios métodos de referencia para el análisis cuantitativo de API y la determinación de otros parámetros químicos y físicos en las muestras utilizadas tanto en la calibración como en la predicción. La preparación de muestras en el laboratorio requiere del uso de la balanza analítica para calcular la concentración de los analitos en las muestras preparadas. Para la sub y sobre dosificación de las muestras, partimos del valor obtenido por el método de referencia para cada preparación farmacéutica, y mediante pesada obtenemos la concentración de API para las nuevas muestras. Otras técnicas analíticas han sido utilizadas:

- HPLC

La determinación del principio activo (Dexketoprofeno) en el producto Enantyum Sobres requirió de un método de referencia HPLC desarrollado en un cromatógrafo Agilent HP 1100 Series HPLC equipado con una columna Hypersil ODS-2 de 12,5cm de longitud, 0,46 cm de diámetro interno y 0.50 μm de tamaño de particular y un detector PDA-UV ($\lambda = 258 \text{ nm}$), utilizando una fase móvil de tampón a pH=3 (65%) y Acetonitrilo (35%). El área de pico para cada analito fue interpolada en una curva de calibración obtenida a partir de soluciones estándar previamente preparadas.

La determinación del principio activo (Dexketoprofeno) y los dos conservantes (metil y propil *p*-hidroxibenzoato) en el producto Fastum Gel requirió de un método de

referencia HPLC desarrollado en un cromatógrafo Agilent HP 1100 Series HPLC equipado con una columna Spherisorb S5 ODS2 de 25 cm de longitud, 0,2 cm de diámetro interno y 10 μm de tamaño de partícula y un detector PDA-UV ($\lambda = 258 \text{ nm}$), utilizando una fase móvil de tampón a pH=3 (65%) y Acetonitrilo (35%). El área de pico para cada analito fue interpolada en una curva de calibración obtenida a partir de soluciones estándar previamente preparadas.

- Espectroscopía UV-Visible.

El contenido de API (Dexketoprofeno) en el producto Enantyum Comprimidos fue determinado a partir de 0,25 g de muestra triturada añadidos a 90 mL de agua Milli-Q. Se aplica un baño de ultrasonidos a la solución durante 15 minutos a 37 °C. Una alícuota de 2 mL es diluida a 5 mL de una solución Metanol/Agua 3:1 de la cual se registra el espectro UV-Vis. El contenido de API se calcula mediante una regresión lineal múltiple (MLR) en la región 210-400 nm.

- Método para la determinación del contenido en agua. Pérdida por desecación.

Para la determinación del contenido de agua, todas las muestras han sido sometidas al método LOD (*lost on drying*), consistente colocar la muestra en una estufa de desecación a 115 °C durante 15 minutos. La pérdida de peso de la muestra se atribuye al contenido de agua de la misma.

- Método para la determinación de la distribución tamaño de partícula. *Particle Size Distribution (PSD)*.

Cada una de las muestras es tamizada en 9 tamices que cubren un rango de tamaños de partícula de 1000 – 38 micras; se calcularon los porcentajes de muestra retenida en cada uno de los tamices y finalmente se obtuvo la curva de distribución granulométrica (PSD) para cada muestra. Con el objeto de construir modelos de calibración adecuados para el seguimiento del proceso, se han seleccionado diferentes intervalos de tamaño de

partícula: inferior a 125 micras, entre 125 y 1000 micras, y superior a 1000 micras. Esta selección se realizó teniendo en cuenta las necesidades del proceso real y seleccionado por la empresa como óptima para el estudio.

- Método para la determinación de la Densidad en muestras granuladas.

La densidad aparente depende de la naturaleza de la materia prima y también del tamaño de partícula de los gránulos que se van formando durante el proceso y tiene elevada influencia en el transporte y manejo del producto final. El valor de densidad se obtuvo pesando un volumen de muestra contenido en una probeta sin compactar ni golpear.

4.3 METODOLOGÍA DEL TRATAMIENTO DE DATOS.

4.3.1 SOFTWARE

A continuación se describe el software utilizado en esta memoria para la adquisición de los espectros NIR.

Adquisición de datos

VISION v2.51 (FOSS NIRSystems, Silver Spring, USA). Programa diseñado para el control del espectrofotómetro NIR FOSS. Permite el registro de los espectros así como su visualización, aplicando los pretratamientos espectrales más habituales. Su entorno permite diseñar rutinas de análisis sencillas para el análisis tanto cualitativo como cuantitativo.

Indico Pro. Permite la adquisición de los espectros registrados con el espectrofotómetro NIR QualitySpec de ASD. Permite registrar espectros individuales y utilizar metódicas para el registro en continuo de los espectros, permitiendo automatizar la monitorización de procesos con varias horas de duración.

TS-Gui. La adquisición de las imágenes ha sido realizada con el software propio de la cámara hiperespectral *TS_Capture* (Think Spectrally, Valencia, España) y el calibrado de la imagen se realiza con una *graphical user interface* (GUI) para MatLab, propio también de la cámara hiperespectral *TS_Gui*.

Tratamiento de datos

MODDE v6.0 (*Umetrics AB, Umea, Sweden*). Permite realizar diseños de experimentos, que han sido utilizados en esta tesis para el diseño de la preparación de muestras en el laboratorio. Además, permite diseñar un conjunto óptimo de experimentos que recoge toda la variabilidad del sistema a estudiar. Una vez realizados los experimentos, permite visualizar las variables más significativas.

The Unscrambler (*Camo Process SA, Trondheim, Norway*). Permite trabajar con conjuntos de datos multivariantes con la posibilidad de utilizar pretratamientos de los datos aplicando SNV (Standard Normal Variate), Mean Centering, correcciones de línea base y derivadas (usando el algoritmo de Savitzky-Golay). Además permite aplicar algunas de las herramientas quimiométricas más habituales para análisis de clasificación (PCA) y para el análisis cuantitativo (PLS).

MatLab v7.0 (*The MathWorks, Massachusetts, USA*). Permite trabajar con datos multivariantes. A través de diferentes rutinas creadas por autores se han utilizado: PLS Toolbox 4.5 (Eigenvector Research, WA, USA) para la aplicación de los algoritmos de PCA y CLS; ALS 2004 de Roma Tauler para la aplicación de MCR-ALS y Augmented MCR-ALS; finalmente, el Statistical Toolbox y Fuzzy Logic toolbox de Mathworks, para la aplicación de algoritmos de clusters, así como el cálculo de los índices de entropía y Silhouette.

BACRA imaging (creado por investigadores del grupo). Estructurado en varias rutinas en Matlab que permite el tratamiento de imágenes des de el preprocesado inicial, con el pretratamiento de los datos, hasta la aplicación de diferentes herramientas quimiométricas para el análisis exploratorio inicial (PCA, coeficientes de correlación,...).

BIBLIOGRAFIA

- [1] Guidance for Industry. PAT—A framework for innovative pharmaceutical development, manufacturing, and quality Assurance. US Department of Health and Human Services, Food and Drug Administration, Center of Drug Evaluation (CDER), and Research, Center for Veterinary Medicine (CVM), Office of Regulatory Affairs (ORA), 2004.
- [2] S. Kueppers, M. Haider. Process analytical chemistry—future trends in industry. *Anal. Bioanal. Chem.* 376:313–315 (2003).
- [3] P. Chalus, Y. Roggo, S. Walter, M. Ulmschneider. Nearinfrared determination of active substance content in intact lowdosage tablets. *Talanta.* 66:1294–1302 (2005).
- [4] M. Blanco, and M. Alcalá. Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy A contribution to process analytical technologies. *Anal. Chim. Acta.* 557:353–359 (2006).
- [5] M. Blanco, M. Alcalá, J. M. González, E. Torras. A process analytical technology approach based on near infrared spectroscopy: tablet hardness, content uniformity, and dissolution test measurements of intact tablets. *J. Pharm. Sci.* 95:2137–2144 (2006).
- [6] C. Peroza, M. A. Santos, and R. J. Románach. Quantitation of drug content in a low dosage formulation by transmission near infrared spectroscopy. *AAPS Pharm. Sci. Tech.* 7(1) (2006), Article 29.
- [7] Herschel W., *Phyl.*, *Trans. Roy. Soc.*, 1800, Part II, 284-292.
- [8] W. Abney, E.R. Festing. *Phyl.*, *Trans. Roy. Soc.*, 1881, 172, 887.
- [9] W. Kaye, *Spectrochimia Acta*, 1954, 6, 257.
- [10] K.H. Norris, J.R. Hart, “Direct spectrophotometric determination of moisture content of grain and seeds” *Proceedings of the 1963 Intl. Symp. on Humidity and Moisture*, Washington D.C., 1965, 4, 19-25.
- [11] K.H. Norris, J.R. Hart, C. Golumbie. “Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extracts”, *Cereal Chemistry*, 1962, 39, 94-99.
- [12] D. Noble. “Analyzing on the INTERNET: Finding analytical chemistry on the Internet is still a challenge, but new resources and search tools are coming on line”, *Anal. Chem.*, 1995, 67(7), 255A-259A.
- [13] C.M. Henry. “Near-IR gets the job done”, *Anal. Chem.*, 1999, 71, 625A-628A.
- [14] B. Jiang, Y. Dong Huang (2008) *J Reinf Plast Comp* 26:1625–1636
- [15] S. Macho, M.S. Larrechi (2002) *Trends Anal Chem* 21:799–806
- [16] M.E. Lafargue, M.H. Feinberg, J.J. Daudin, D.N. Rutledge. (2003) *Anal Bioanal Chem* 375:496–504
- [17] M. Blanco, M. Bautista, M. Alcalá. (2008) *J Pharm Sci* 97:1236–1245
- [18] J.L. Jensen. “Acceptance of NIR as an analytical method”, *Process Control and Qual.* 1997, 9, 161-166.
- [19] A.M.C Davies, A. Garrido. “The Proceeding of NIR-2003”, 2004, Ed. NIR publications.
- [20] T. Davies. “The history of near infrared spectroscopic analysis: past, present and future from sleeping technique to the morning star of spectroscopy”, *Analisis*, 1998, 26(4) 17-19.
- [21] E.N. Lewis, J.E. Carroll, and F.C. Clarke, *NIR News* 12(3), 16-18 (2001)
- [22] D.A. Skoog, F.J. Holler, T.A. Nieman. (2001) *Principios de análisis instrumental*, 5ª ed., McGraw- Hill.
- [23] J.J. Workman Jr. (1999) Review of process and non-invasive near infrared and infrared spectroscopy: 1993-1999. *Appl. Spectrosc. Reviews* 34, 1-89.
- [24] R.F. Goddu, D.A. Delker. (1960) Spectra-Structure Correlations for Near-Infrared Region. *Anal. Chem.*, 32, 140-141.
- [25] E.W. Ciurczak. (2001) Principles of Near-Infrared Spectroscopy. In Burns, D.A. and Ciurczak, E.W. (eds), *Handbook of Near-Infrared Analysis*, 2nd Edn; Marcel Dekker, Inc.; New York, pp 9-18.

BIBLIOGRAFIA

- [26] A.S. Bonanno, P.R. Griffiths. "Discrimination of organic solvents using an infrared-emitting diode-based analyzer. Part I: Feasibility", *Appl. Spectrosc.*, 1995, 49(11), 1590-1597.
- [27] United States Patent. Patent Number 4.286.327, 1981.
- [28] D.A. Skoog. "Principles of Instrumental Analysis", 1997, Ed. Brooks Cole.
- [29] M. Blanco, M. Alcalà. (2006). Simultaneous quantitation of five active principles in a pharmaceutical preparation: Development and validation of a near infrared spectroscopic method. *Eur J Pharm Sci* 27:280–286.
- [30] J. Colón, C. Peroza, W. Caraballo, C. Conde, T. Li, K.R. Morris, R.J. Romañach. (2005). On line non-destructive determination of drug content in moving tablets using near infrared spectroscopy. *Proc Anal Tech* 2:8–15.
- [31]. I. Paris, A. Janoly-Dumenil, A. Paci, L. Mercier, P. Bourget, F. Brion, P. Chaminade, A. Rieutord. (2006). Near infrared spectroscopy and process analytical technology to master the process of busulfan paediatric capsules in a university hospital. *J Pharmaceut Biomed Anal* 41:1171–1178.
- [32] M. Blanco, M.A. Romero. (2001) Near-infrared libraries in the pharmaceutical industry: a solution for identity confirmation, *Analyst* 126, 2212-2217.
- [33] P.L. Walling, J.M. Dabney. Moisture in skin by nearinfrared reflectance spectroscopy. *J. Soc. Cosmet. Chem.*, 40, 151-171 (1989).
- [34] K. Martin. Direct measurement of moisture in skin by NIR spectroscopy. *J. Soc. Cosmet. Chem.*, 44, 249-261 (1993).
- [35] K. Martin. *In vivo* measurements of water in skin by nearinfrared reflectance. *Appl Spectrosc.*, 52, 1001-1007 (1998).
- [36] M.C. Pasikatan, J.L. Steele, C.K. Spillman, E. Haque, *Journal Near Infrared Spectroscopy* 9 (1998) 153.
- [37] P. Frake, I. Gill, C. N. Luscombe, D. R. Rudd, J. Waterhouse and U. A. Jayasooriya. Near-infrared mass median particle size determination of lactose monohydrate, evaluating several chemometric approaches *Analyst*, 1998, 123, 2043–2046
- [38] H.G. Britain (1999) *Polymorphishm in Pharmaceutical Solids*, Ed by H.G. Britain, Marcel Dekker Inc.
- [39] M. Blanco, D. Valdés, I. Llorente, M. Bayod. (2005) Application of NIR spectroscopy in polymorphic analysis: Study of pseudo-polymorphs stability. *J. Pharm. Sci.* 94(6), 1336-1342.
- [40] M. Blanco, M. Alcalà, J.M. González, E. Torras. 2006. Near infrared spectroscopy in the study of polymorphic transformations. *Anal Chim Acta* 567: 262–268.
- [41] B. D. Mindel, *Process Control and Quality*, 9, 173, 1997.
- [42] M. Blanco y D. Serrano, *Analyst*, 123, 2059-2064, 2000.
- [43] J. A. Westerhuis, S. P. Gurden, A. K. Smilde, *Anal. Chem.*, 72, 5322-5330, 2000.
- [44] C. Coffey, B. E. Cooley, D. S. Walker, *Anal. Chim. Acta*, 395, 335-341, 1999.
- [45] R. A. Heikka, et. al, *Anal. Chim. Acta*, 349, 287-294, 1997.
- [46] O. Berntsson, L.G. Danielsson, B. Lagerholm, S. Folestad, *Powder Technology* 123 (2002) 185–193.
- [47] S.S. Sekulic, H.W. Ward, D.R. Brannegan, E.D. Stanley, C.L. Evans, S.T. Sciavolino, P.A. Falley and P.K. Aldridge, *Anal. Chem*, 68 (1996) 509–513.
- [48] A.C. Jorgensen, J. Rantanen, P. Luukkonen, S. Laine, J. Yliruusi. *Anal. Chem.* 76 (2004) 5331–5333
- [49] A.M. Railkar, J.B. Schwartz. Evaluation and Comparison of a Moist Granulation Technique to Conventional Methods. *Drug Development and Industrial Pharmacy* 26[8], 885-889. 2000.
- [50] A. Faure, P. York, R. Rowe. Process control and scale-up of a pharmaceutical wet granulation processes: a review. *European Journal of Pharmaceutics and Biopharmaceutics* 52, 269-277. 2001.
- [51] J. Rantanen, A. Jorgensen, E. Räsänen, P. Luukonene, S. Airaksinen, J. Raiman, K. Hänninen, O. Antikainen, J. Yliruusi. Process Analysis of Fluidized Bed Granulation. *AAPS PharmSci.*; 2(4) (2001)

BIBLIOGRAFIA

- [52] M. Otsuka, Y. Mouri, Y. Matsuda. Chemometric Evaluation of Pharmaceutical Properties of Antipyrine Granules by Near-Infrared Spectroscopy. *AAPS PharmSci.*; (4) 3. (2003)
- [53] J. Rantanen, H. Wikström, R. Turner, L.S. Taylor. Use of In-Line Near-Infrared Spectroscopy in combination with chemometrics for improved Understanding of Pharmaceutical Processes. *Anal. Chem.* 77, 556-563. (2005)
- [54] F.J.S Nieuwmeijer, M. Damen, A. Gerich, F. Rusmini, K.V. Maarschalk, H. Vromans, Granule Characterization During Fluid Bed Drying by Development of a Near Infrared Method to Determine Water Content and Median Granule Size. *Pharm. Res.*, 24 (2007).
- [55] B. Rambali, L. Baert, D.L. Massart. Scaling up of the fluidized bed granulation process. *Int. J. Pharm* 252. 197-206. (2003)
- [56] B. Rambali, L. Baert, D.L. Massart. Using experimental design to optimize the process parameters in fluidized bed granulation on a semi-full scale. *Int. J. Pharm.* 220 149-160. (2001)
- [57] D. Vojnovic, D. Chicco, H. El Zenary. Doehlert experimental design applied to optimization and quality control of a granulation process in a high shear mixer. *Int. J. Pharm.* 145 203-213. (1996)
- [58] F. Koehler, E. Lee, L.H Kidder, E.N. Lewis. Near infrared spectroscopy: the practical chemical imaging solution. *Spectroscopy Europe* (2002), 14(3), 12-14, 16, 18-19.
- [59] A.S. El-Hagrasy, H.R. Morris, F. D'Amico, R.A. Lodder, J.K. Drennen III. Near-infrared spectroscopy and imaging for the monitoring of powder blend homogeneity, *J. Pharmaceutical Science* 90 (2001) 1298-1307.
- [60] E. N. Lewis, J. E. Carroll, F. M. Clarke, NIR Imaging: A Near Infrared View of Pharmaceutical Formulation Analysis. *NIR News* 12, 2001, 16.
- [61] F. Clarke. Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. *Vibrational Spectroscopy* 34 (2004) 25-35.
- [62] G. Reich. Near-Infrared spectroscopy and Imaging: Basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews* 57 (2005) 1109-1143.
- [63] S. Wold. "Spline Functions, a new tool in data-analysis", *Kemisk Tidskrift*, 1972, 84(3), 34-37
- [64] D. L. Massart, et. al, *Chemometrics: a textbook*, Elsevier, 1988.
- [65] R. Kramer, *Techniques for quantitative analysis*, Marcel Dekker, 1998.
- [66] H. L. Martens y T. Naes, *Multivariate calibrations*, Wiley, 1989.
- [67] C.E. Miller. (2005) Chemometrics in Process Analytical Chemistry. In: K. Bakeev (ed) *Process Analytical Technology*. Blackwell publishing, Oxford, pp 226-227.
- [68] D.L. Massart, B.G.M Vandeginste, L.M.C. Buydens, S.P. De Jong, J. Lewi. and Smeyers-Verbeke, S.(1997) In: B.G.M Vandeginste and S.C Rutan, *Handbook of Chemometrics and Qualimetrics: Part A*, Elsevier pp. 1-20.
- [69] M. Blanco, M. Alcalá. (2005) Use of Near-Infrared Spectroscopy for Off-line Measurements in the Pharmaceutical Industry. In: Bakeev, K. (ed) *Process Analytical Technology*. Blackwell publishing, Oxford, pp 362-391.
- [70] R. J. Barnes, M. S. Dhanoa y S. J. Lister, *Appl. Spectrosc.*, 113, 1849, 1988.
- [71] W. Plugge y C. Van der Vlies, *J. Pharm. Biomed. Anal.*, 14, 891, 1996.
- [72] A. Savitzky y M. J. E. Golay, *Anal. Chem.*, 36, 1627, 1964.
- [73] K.L.A. Chan, N. Elkhider, S.G. Kazarian, *Chem. Eng. Res. Des.* 83(2005) 1303.
- [74] Martens, H. and Næs, T. (1991) *Multivariate Calibration*; Wiley; New York.
- [75] R. Tauler, A. Smilde and B.R. Kowalski. Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution. *Journal of Chemometrics*, 1995, 9, 31 - 58
- [76] R. Tauler, B.R. Kowalski and S. Flemming. Multivariate Curve resolution Applied to Spectral Data from Multiple Runs of an Industrial Process. *Analytical Chemistry*, 1993, 65, 2040 - 2047
- [77] R. Tauler, *Anàlisi de mesclades mitjançant resolució multivariant de corbes*, Institut d'estudis catalans, 1997.

BIBLIOGRAFIA

- [78] A. de Juan, R. Tauler, *Crit. Rev. Anal. Chem.* 36 (2006) 163
- [79] L. Zhang, M.J. Henson, S.S. Sekulic, *Anal. Chim. Acta* 545 (2005) 262.
- [80] B. Walczak, *Wavelets in Chemistry*, Elsevier, New York, USA, 2000.
- [81] L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, Chichester, West Sussex, UK, 1990.
- [82] S. Theodoridis, K. Koutroumbas, *Pattern Recognition*, Academic Press, New York, USA, 1999.
- [83] J.B. MacQueen, *Some Methods for Classification and Analysis of Multivariate Observations*, University of California Press, Berkeley, CA, USA, 1967.
- [84] N. Jovanovic; A. Gerich, A. Bouchard, W. Jiskoot Near-infrared imaging for studying homogeneity of protein-sugar mixtures. *Pharmaceutical research*(2006), 23(9), 2002-13.
- [85] M. Blanco, J. Cruz, M. Bautista. Development of a univariate calibration model for pharmaceutical analysis based on NIR spectra. *Analytical and Bioanalytical chemistry*. 2008
- [86] AG. Asuero, A. Sayago, AG. Gonzalez (2006) *Crit Rev Anal Chem* 36:41–59
- [87] H. Mark. (2001) Qualitative discriminant analysis. In Burns, D. and Ciurczak, E.W. (eds); *Handbook of Near-Infrared Analysis*, 2nd Edition; Marcel Dekker; New York, pp. 351–362.
- [88] A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, *Trends Anal. Chem.* 23 (2004) 70.

5. RESULTADOS Y DISCUSIÓN

5. RESULTADOS Y DISCUSIÓN

A continuación se recogen los resultados más relevantes divididos en función del tipo de información obtenida. Los resultados han sido distribuidos en función del tipo de muestra estudiada. Se han clasificado en: 1) muestras de producto final del proceso farmacéutico, que han sido analizadas en el laboratorio y de las cuales han sido estudiados parámetros químicos y físicos, 2) muestras durante la monitorización de un proceso de granulación, de las que han sido de nuevo analizados parámetros químicos y físicos, y finalmente 3) se muestran los resultados con muestras de laboratorio mediante el uso de técnicas de NIR Imagen, y cuyos resultados se han clasificado en función de la información obtenida según la herramienta quimiométrica utilizada.

5.1 DETERMINACIÓN DE PARÁMETROS DE INTERÉS EN PRODUCTO ACABADO

Los primeros resultados mostrados corresponden a estudios realizados sobre producto final de 3 preparados farmacéuticos. Previo a la determinación de los parámetros de interés, debemos identificar previamente el producto objeto de análisis. Con este objetivo se han realizado bibliotecas espectrales que permiten rechazar muestras que presentan diferencias con el producto teórico.

Una vez identificado el producto se puede proceder a la determinación de API; en uno de los productos finales se ha estudiado la posible transformación polimórfica del API durante el proceso de producción.

IDENTIFICACIÓN DE PRODUCTOS FARMACÉUTICOS

El objetivo de las bibliotecas espectrales es identificar muestras farmacéuticas de producto antes de realizar las determinaciones químicas y físicas necesarias para el control de calidad.

RESULTADOS Y DISCUSIÓN

El valor umbral de correlación (*threshold*) se fijó de tal manera que una muestra de producto acabado se identificara correctamente y a la vez que la biblioteca fuera capaz de diferenciarla del resto de los compuestos que forman parte del preparado farmacéutico.

Las tablas 5.1 y 5.2 muestran la biblioteca espectral creada para el producto Enantyum comprimidos, y tiene como objetivo discriminar el producto final, un comprimido recubierto, de todos los excipientes y API, y también de comprimidos sin recubrir.

La biblioteca de referencia creada para la correcta identificación del preparado farmacéutico contiene espectros de lotes de las materias primas (excipientes y principio activo) recogidos durante un periodo de 2 años mientras que los lotes de comprimidos lacados y no lacados corresponden a un periodo de 5 meses. La biblioteca de referencia ha sido desarrollada aplicando la 2ª derivada de Norris con un segmento de 10 puntos y un gap de 10 puntos y estableciendo como valor umbral un coeficiente de correlación 0.97. Las siguientes tablas muestran los coeficientes de correlación entre todos los excipientes, API, y comprimido lacado y no lacado, con el espectro del comprimido lacado.

Tabla 5.1. Rangos de los coeficientes de correlación de la biblioteca de identificación. Se han representado los coeficientes de correlación entre los espectros de todas las materias primas y producto final con el espectro del comprimido lacado.

Coefficiente de correlación	
	Comprimido lacado
Comprimido lacado	1
Comprimido no lacado	0.991
Almidón de maíz	0.758
Celulosa microcristalina ph101	0.859
Carboximetilalmidón	0.700
Dexketoprofeno Trometamol	0.303
Diesterato Glicerol	0.332

RESULTADOS Y DISCUSIÓN

Debido a las ambigüedades existentes entre las dos clases de comprimidos, con un coeficiente de correlación entre ambos de 0.991, es necesaria la cualificación de dichos productos para poder clasificar correctamente comprimidos lacados y no lacados. El método de cualificación utilizado es la distancia de Mahalanobis.

Tabla 5.2. Parámetros de la biblioteca de cualificación mediante distancia de Mahalanobis.

	Método	Rango (nm)	Pretratamiento	Nº PC	Threshold
Comprimido lacado	Distancia Mahalanobis	1110-2188	2da Derivada	3	0.46
Comprimido núcleo	Distancia Mahalanobis	1110-2188	2da Derivada	3	0.46

Aplicando estos parámetros a la biblioteca espectral (tabla 5.1), se han podido clasificar correctamente los comprimidos lacados de los no lacados, cumpliendo el primer objetivo del estudio para este preparado farmacéutico, que tiene como etapa final, una vez identificado correctamente el producto, determinar el contenido de API presente en el mismo, y que se muestra en el próximo apartado.

Esta misma estrategia de desarrollo de bibliotecas de identificación se ha utilizado como paso previo a los métodos de análisis cuantitativos. Todas las bibliotecas espectrales desarrolladas permiten identificar correctamente el producto acabado, sin mostrar confusión con el resto de componentes puros.

DETERMINACIÓN DE API

En la presente memoria se ha determinado el contenido de API en 3 productos farmacéuticos. Cada uno de estos productos presenta un formato diferente: comprimido, gel y granulado, y cada uno de los casos se han abordado utilizando metodologías diferentes para la calibración. Para confirmar la bondad de los modelos

RESULTADOS Y DISCUSIÓN

construidos, los resultados obtenidos han sido comparados con los valores obtenidos para cada uno de los métodos de referencia de cada producto.

Cuantificación de Dexketoprofeno en comprimidos farmacéuticos.

El estudio de este producto farmacéutico se dividió en 2 etapas: en primer lugar la correcta identificación del producto final (apartado anterior) y la determinación del contenido de API en el producto final.

El producto intermedio corresponde a un granulado previo antes de la compactación final en comprimidos y posterior recubrimiento (lacado)

Se han desarrollado dos modelos de calibración: para muestras granuladas al final del proceso de granulación (modelo 1) y para comprimidos recubiertos en el control de calidad para la liberación del producto (modelo 2). Para la construcción de los dos modelos se usan diferentes formatos de la misma muestra: comprimidos lacados molturados y comprimidos lacados intactos.

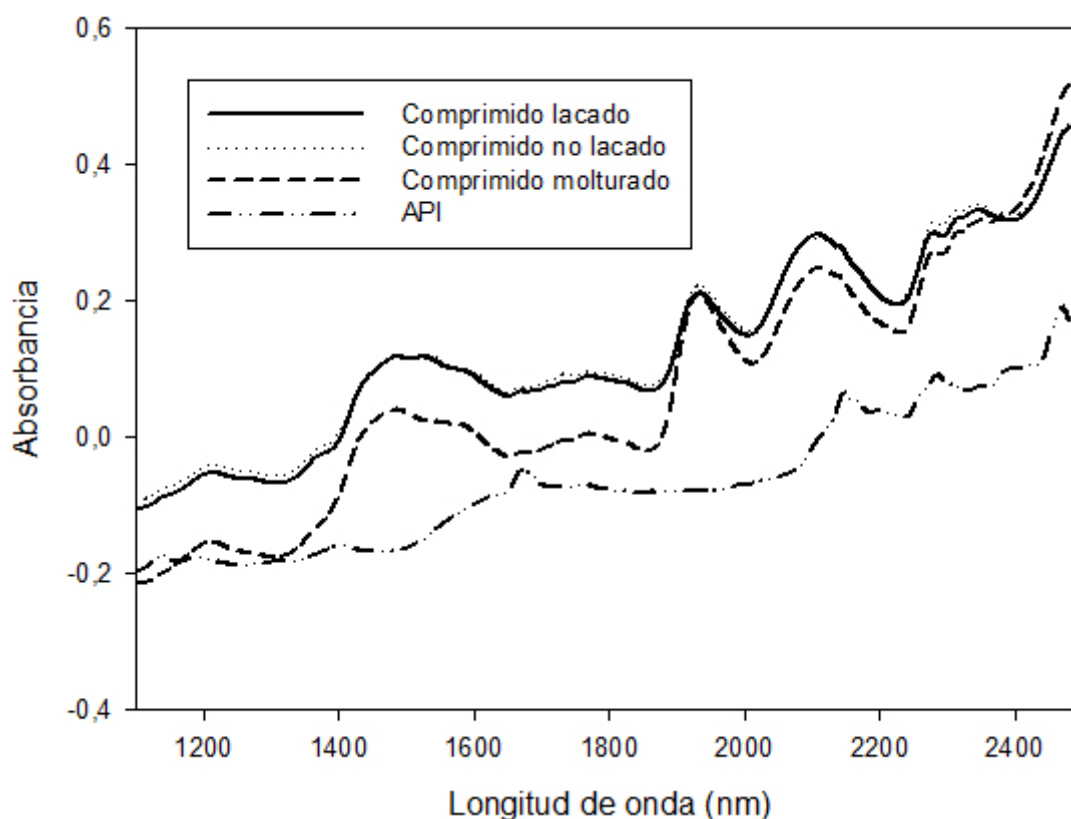


Figura 5.1. Espectros en Absorbancia del API, la misma muestra presentada en : comprimido lacado, comprimido no lacado y comprimido molturado.

RESULTADOS Y DISCUSIÓN

La Figura 5.1 muestra el espectro en modo Absorbancia del API y de la misma muestra de producción: comprimido no lacado, comprimido lacado y el mismo comprimido lacado molturado; los tres espectros muestran el mismo perfil debido a que no hay variación en la composición química; sin embargo se observa un desplazamiento de los espectros debido a las diferencias físicas que presentan las muestras. La única diferencia entre los comprimidos lacados y no lacados es el recubrimiento de laca, que además de estar en muy baja concentración, no presenta bandas de absorción significativas. Los comprimidos presentan los valores más elevados en absorbancia respecto al espectro de la misma muestra granulada (desplazamiento de la línea base) debido al efecto de la compresión; a mayor compresión el espectro presenta un mayor desplazamiento. Entre los dos tipos de comprimidos, no se observan diferencias significativas entre los correspondientes espectros, ya que lo único que los diferencia es la capa de recubrimiento.

Para la construcción del modelo 1 se utilizaron muestras procedentes de la pulverización de comprimidos lacados junto con otras muestras preparadas por sobre e infra dosificación.

Con estas muestras se ensayaron diferentes modelos, aplicando diferentes pretratamientos y rangos espectrales. La Tabla 5.3 muestra los parámetros más significativos del modelo para muestras granuladas y las predicciones obtenidas con muestras granuladas de producción; el pretratamiento que mejores resultados aportó fue el de 2º Derivada de Savitzky-Golay. El modelo fue construido con 20 muestras granuladas, de las cuales 13 eran dosificadas con un intervalo de $\pm 20\%$ de API y las otras 7 contenían el valor nominal de API. Se observa que con tres factores PLS, se obtienen unos estadísticos de calibración correctos (99.1% Varianza explicada) y una adecuada capacidad predictiva (RSEP=1.01%) para muestras granuladas de producción.

RESULTADOS Y DISCUSIÓN

Tabla 5.3. Parámetros significativos de los modelos de calibración para muestra granulada y comprimido final lacado.

Modelo		
	1. Granulado	2. Comprimidos
	(1era etapa)	(2da etapa)
Pretratamiento	2nd der.	2nd der.
Rango espectral (nm)	1134–1798	1134–1798
NºFactores PLS	3	3
Varianza Y %	99.12	99.31
SET CALIBRACIÓN		
Muestras granuladas	20	13
Comprimidos	—	5
Concentración (mg/g)	75.6–120.6	75.6–120.6
RSEC	1.04	0.97
SET PREDICCIÓN		
Muestras Granuladas	20	—
RSEP	1.01	—
Comprimidos	28	28
RSEP	5.03	3.43

Con este modelo se analizaron además, los comprimidos lacados y no lacados. Las predicciones de dichas muestras presentan errores sistemáticos con predicciones siempre por exceso además de un RSEP > 5%, lo que indica que las diferencias generadas por la compactación producen errores significativos y por tanto el modelo 1 no es válido para predecir correctamente comprimidos.

RESULTADOS Y DISCUSIÓN

Se realizó un PCA de las muestras granuladas del modelo 1 y de comprimidos lacados y sin lacar; la figura 5.2 muestra el gráfico de dispersión de un PCA obtenido con espectros en 2ª Derivada de Savitzky-Golay que muestra que el primer factor PC1 (65% Varianza) recoge esencialmente las diferencias físicas de los grupos de muestras.

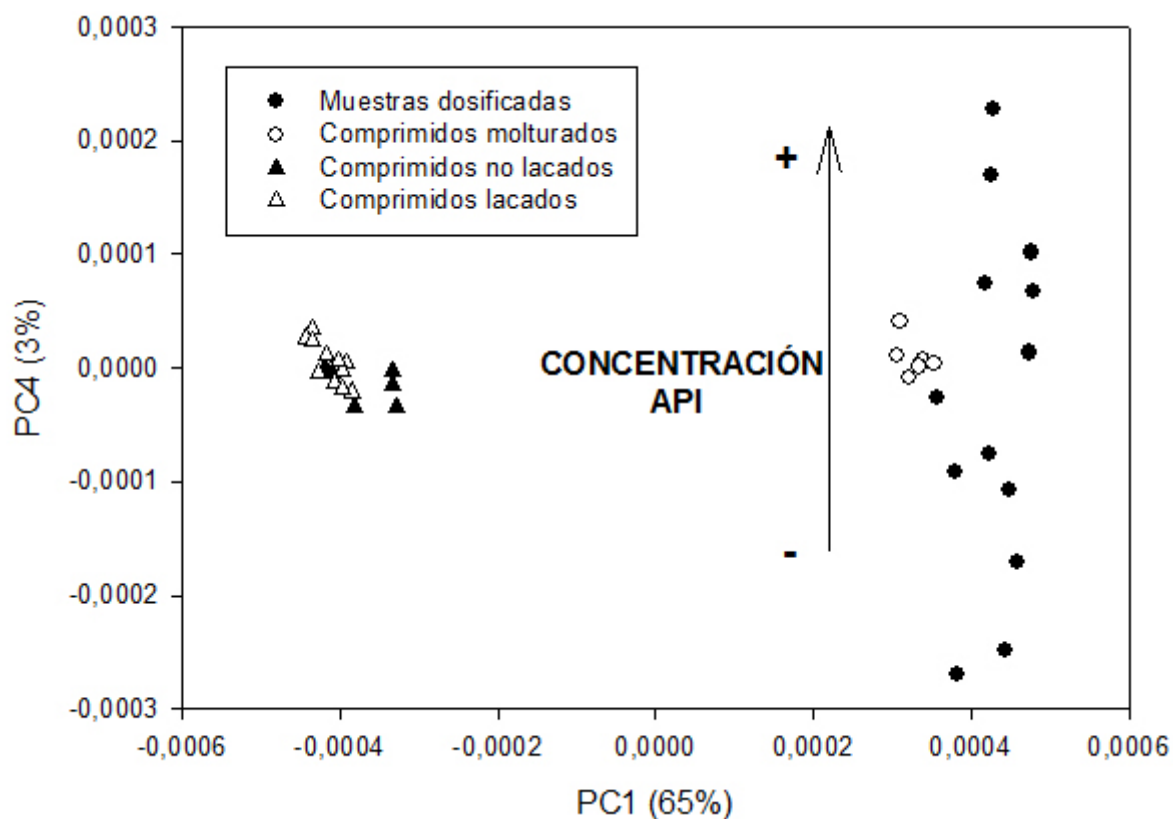


Figura 5.2. Gráfico de dispersión de los scores al realizar un PCA con los espectros de Absorbancia de muestras granuladas en el laboratorio y muestras de producción.

A la derecha del gráfico se sitúan las muestras molturadas, mientras que los comprimidos se han agrupado a la izquierda.

Además se observó que es el PC4, el componente principal que recoge la variabilidad química de las muestras. Los comprimidos lacados y no lacados, y los comprimidos de producción molturados (con prácticamente el mismo contenido de API) están situados en el centro del eje del PC4, mientras que las muestras dosificadas están distribuidas a lo largo del eje según la concentración de API presente.

RESULTADOS Y DISCUSIÓN

La formación de dos conjuntos de muestras en el gráfico según el formato de presentación del producto explica la incorrecta predicción de los comprimidos al utilizar el modelo 1 para granulados.

Para construir un modelo para los comprimidos lacados se requirió de muestras que cubrieran un amplio rango de concentraciones. Ante la imposibilidad de ampliar el rango de concentración de API en comprimidos recubiertos (estos ya no se pueden modificar), se utilizaron el conjunto de muestras preparadas para ampliar el rango de API del modelo 1 que incorporan la variabilidad química en el modelo. La variabilidad física, debida principalmente a la compactación de la muestra granulada, fue incluida en el modelo incorporando comprimidos recubiertos de producción. Aunque la adición de comprimidos recubiertos al modelo introduce la fuente de variación del recubrimiento lacado, ya se advirtió en la figura 1 que su influencia sobre el espectro es mínima.

La tabla 5.3 muestra los parámetros más significativos del modelo PLS construido para predecir comprimidos lacados del final del proceso productivo (modelo 2). Se ensayaron también diferentes rangos espectrales, pretratamientos y grupos de muestras y los mejores modelos se obtuvieron con los mismos parámetros que el modelo anterior: 3 factores PLS fueron necesarios de nuevo para predecir correctamente los comprimidos lacados de producción.

La predicción de los comprimidos lacados presenta un valor de RSEP de 1.63, disminuyendo considerablemente respecto a la predicción de los mismos con el modelo 1 y los residuales se distribuyen aleatoriamente alrededor del 0, por lo que puede considerarse un modelo adecuado para predecir las muestras del final del proceso.

Comprimidos no lacados fueron evaluados con este modelo obteniendo un valor de RSEP de 2.30 %. La predicción de este tipo de comprimidos produce un error sistemático (figura 5.3 a) por exceso en todos los lotes, atribuible a la falta de recubrimiento.

RESULTADOS Y DISCUSIÓN

Además, fueron evaluadas muestras granuladas con el modelo 2, obteniendo un valor de RSEP de 3.43 % y, al contrario de los comprimidos no recubiertos, con un error sistemático por defecto en todas las muestras (figura 5.3 b).

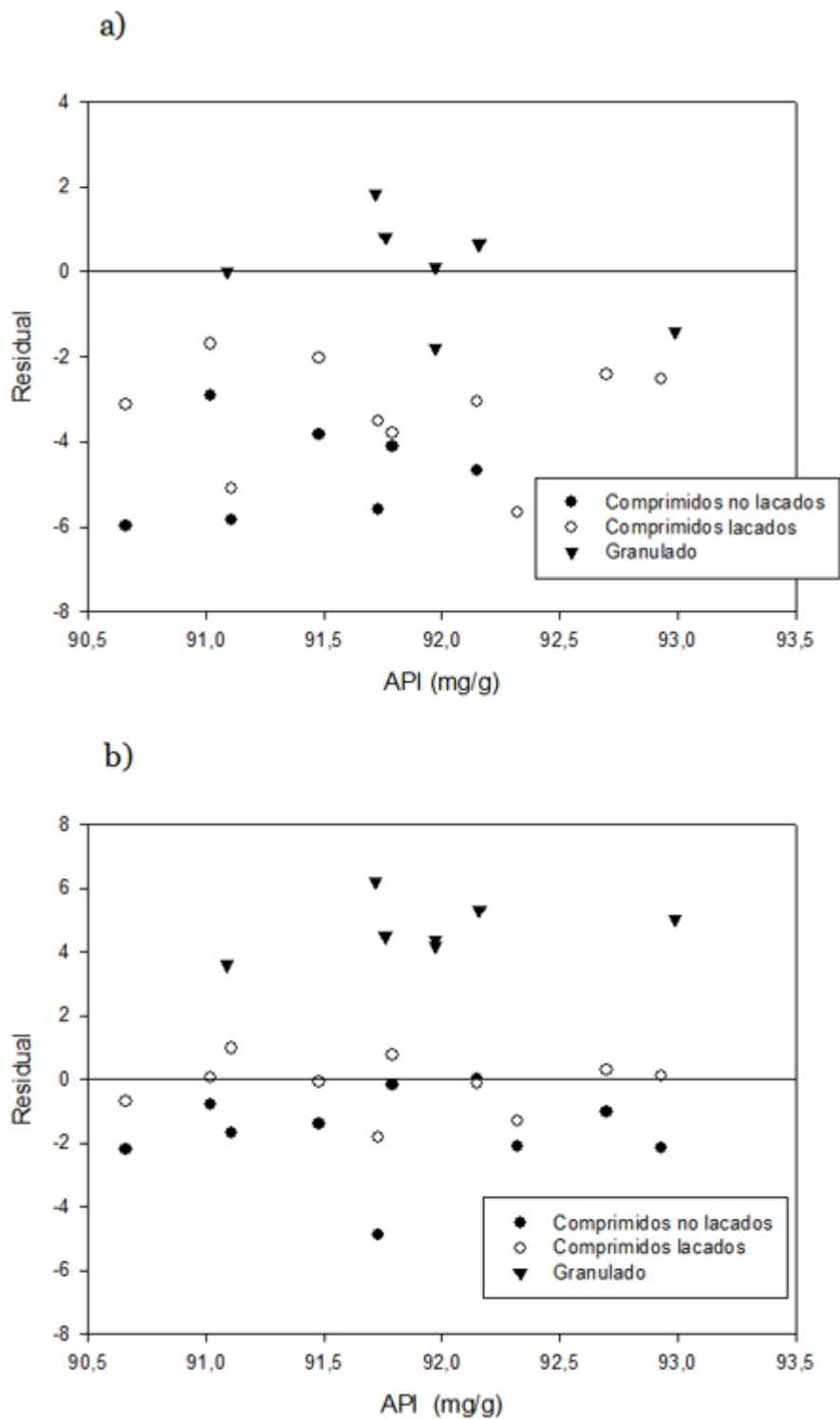


Figura 5.3. Gráfico de residuales al evaluar los diferentes tipo de muestras con el modelo 1 (a) y el modelo 2 (b).

RESULTADOS Y DISCUSIÓN

Los dos modelos construidos presentan parámetros de calibración y predicción óptimos en función de la muestra, por lo que pueden ser utilizados dentro del control de calidad del producto. El modelo 1 puede ser utilizado para determinar el contenido de API en el producto intermedio del proceso (granulado), mientras que el producto final (comprimido lacado) puede ser correctamente identificado (biblioteca espectral) y correctamente cuantificado (modelo 2).

Ante la complejidad de las muestras analizadas y la dificultad de construir un único modelo para las dos etapas, ha sido necesario construir un modelo en función del tipo de muestra, que permiten conocer el contenido de API durante, y al final del proceso de producción. La posibilidad de utilizar el modelo 1 en el control rutinario podría permitir actuar sobre el proceso en caso que fuese necesario añadir API o excipientes para conseguir el valor deseado, evitando llegar al final del proceso, donde un error de dosificación significaría la pérdida de un lote de producción.

Cuantificación de Ketoprofeno en muestras de gel farmacéutico.

Una alternativa al uso de comprimidos antiinflamatorios son las pomadas o geles que son aplicadas directamente sobre la zona de dolor. El análisis de este tipo de muestras mediante Espectroscopia NIR requiere del uso de medidas por transfectancia, donde se ha determinado el contenido de 3 analitos (Ketoprofeno (KET), dos conservantes p-hidroxibenzoato de metilo (MET) y de propilo (PRO)) en un producto que presenta una matriz agua/etanol, con una gran absorbancia en la región del NIR. Una vez construidos los correspondientes modelos para la determinación de los analitos, estos han sido correctamente validados.

La mayor dificultad del análisis de este gel reside en los analitos que están presentes a bajas concentraciones. Si bien el Ketoprofeno presenta una concentración de 2,5%, los dos conservantes están en concentraciones inferiores a 0,1%. Además de la baja concentración, la elevada semejanza química de los 2 conservantes aporta, si cabe, mayor dificultad al análisis.

RESULTADOS Y DISCUSIÓN

Los espectros NIR del preparado farmacéutico, principio activo y sus conservantes se muestran en la figura 6.9. El espectro del preparado farmacéutico se caracteriza por poseer dos bandas anchas e intensas a 1470 y 1910 nm, correspondientes al primer sobretono y a la banda de combinación del enlace OH. El agua es el excipiente mayoritario (se encuentra a una concentración del 54%) y de mayor absorbancia, y consecuentemente sus bandas características son las predominantes en el espectro del preparado farmacéutico. Otras bandas distinguibles aparecen en la zona de 2300 nm correspondientes a la presencia de Etanol (42%). Las bandas más significativas del principio activo (KET) y conservantes (parte superior figura 6.9) se encuentran entre 1500-1800 nm y también entre 2000-2500 nm pero su baja concentración no queda reflejada en el espectro de las muestras.

RESULTADOS Y DISCUSIÓN

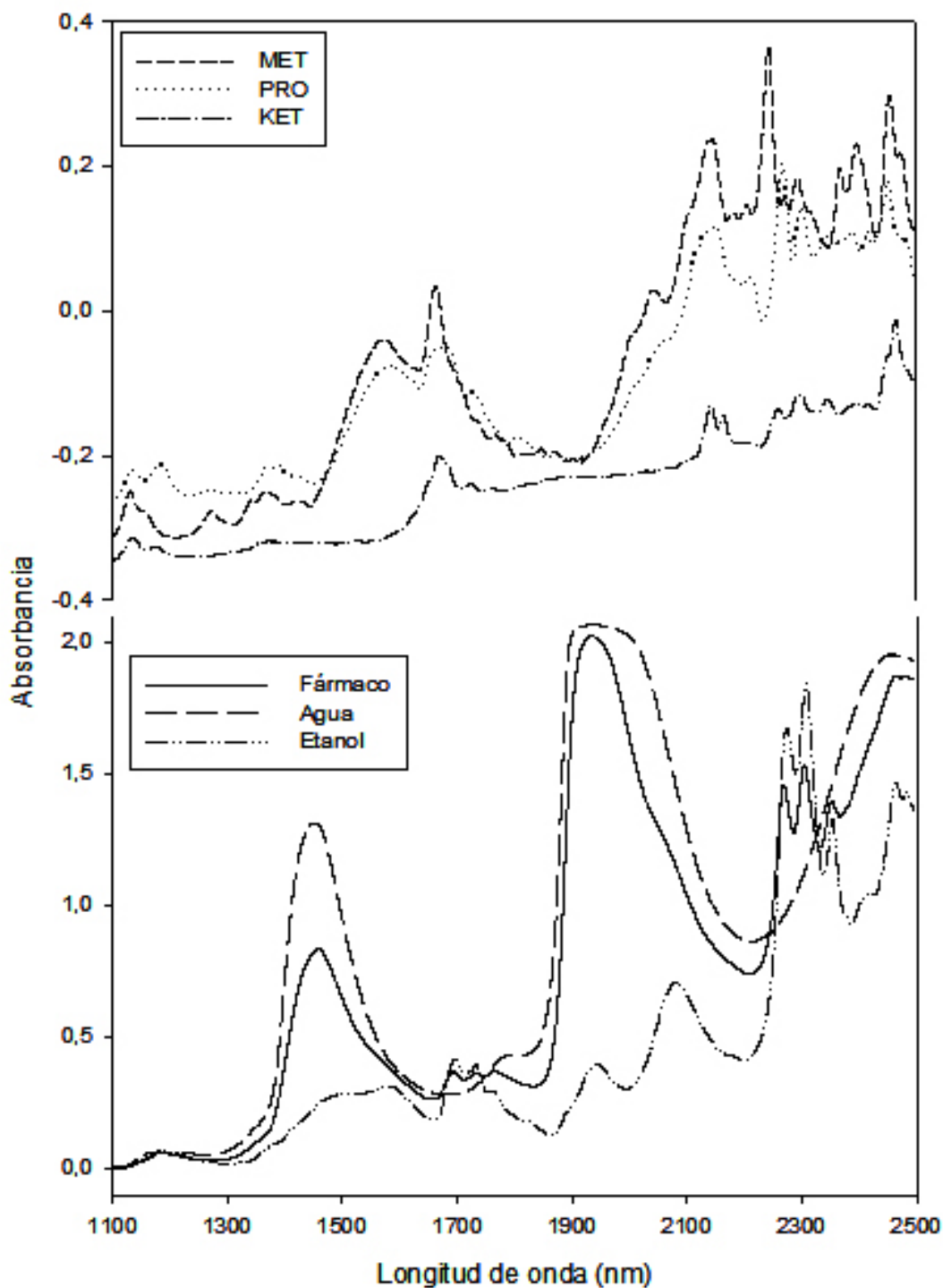


Figura 5.4. Espectro NIR en modo Absorbancia del principio activo (KET), los dos conservantes (MET y PRO), el preparado farmacéutico y los dos excipientes mayoritarios (agua y etanol).

La baja concentración de los conservantes y su elevada semejanza espectral, exigió para el desarrollo de los modelos de calibración un trabajo laborioso de diseño de

RESULTADOS Y DISCUSIÓN

concentraciones y selección de muestras, pretratamiento espectral y selección del intervalo de longitudes de onda. La preparación de las muestras de laboratorio se ha realizado minimizando la correlación entre las concentraciones entre pares de analitos.

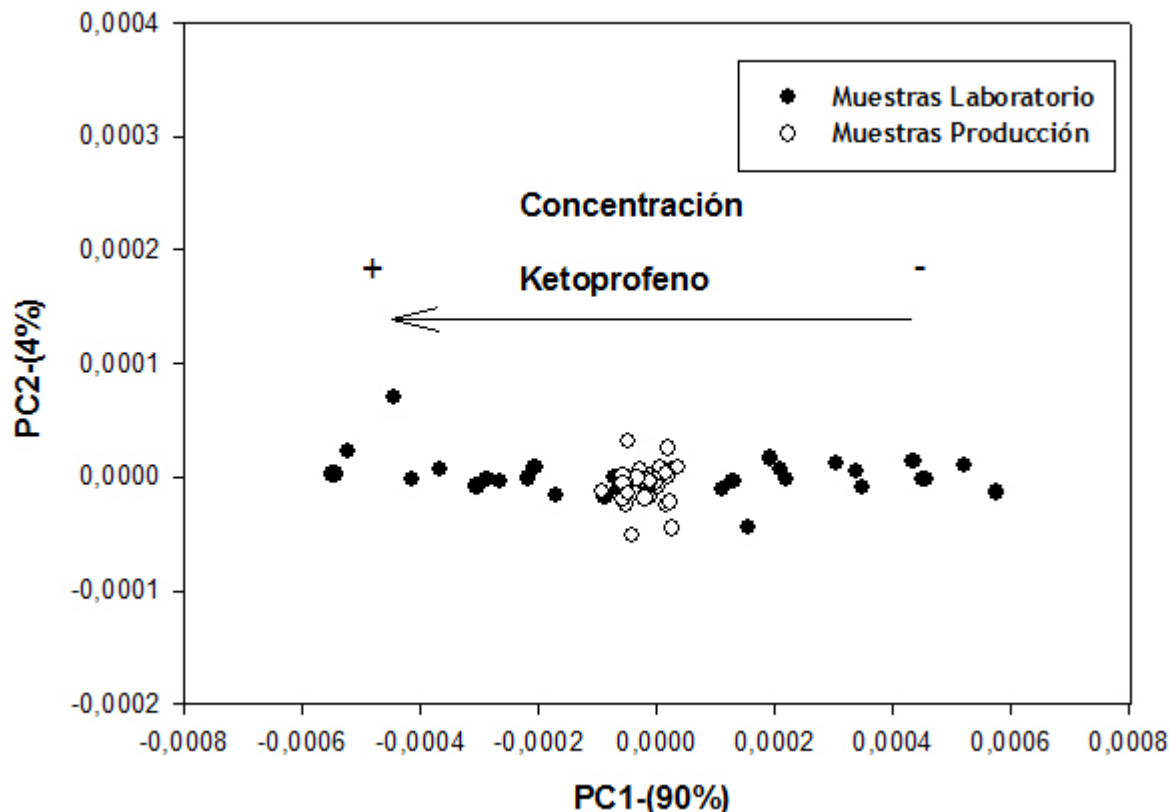


Figura 5.5. Gráfico de dispersión de los scores al realizar un PCA con los espectros en Segunda Derivada de Savitzky-Golay de muestras de laboratorio y muestras de producción.

El primer paso para la construcción del modelo para la cuantificación de los analitos consistió en comparar las muestras preparadas en el laboratorio y las muestras reales de producción. Un gráfico de scores (PC2 vs PC1) correspondiente a un PCA de los espectros en 2ª Derivada de S-G en el intervalo espectral completo (1100-2500nm) de las muestras de producción y de preparadas en el laboratorio indica que la concentración de Ketoprofeno es la mayor fuente de variabilidad en los espectros (Figura 5.5); A diferencia del estudio anterior, donde era necesario introducir muestras de producción junto a las dosificadas para construir un modelo robusto, las muestras

RESULTADOS Y DISCUSIÓN

preparadas en el laboratorio y las muestras de producción presentan como fuente de variabilidad más relevante la concentración de API, por lo que no es necesaria la inclusión de muestras de producción en el conjunto de calibración para la construcción del modelo de calibración del Ketoprofeno. En este preparado farmacéutico y a diferencia de las muestras sólidas, el efecto de dispersión (scattering) producido por el tamaño de partícula no se observa en este tipo de muestras.

Los parámetros más significativos de los modelos construidos para la determinación de los tres analitos se muestran en la Tabla 5.4. El pretratamiento espectral que mejores resultados ofreció fue la 2ª derivada de Savitzky-Golay en todos los casos. El criterio de Jack-Knifing sugirió la utilización de los intervalos espectrales citados en la tabla. Si comparamos estos intervalos con los espectros de los analitos puros de la Figura 5.4, observamos como la selección espectral coincide con las bandas más significativas de los tres analitos. El número de factores PLS necesario para establecer el modelo aumenta a medida que disminuye la concentración del analito, siendo necesarios hasta 6 factores para el PRO.

Tabla 5.4. Parámetros de los modelos de calibración construidos.

	KET	MET	PRO
Pretratamiento	2da Der	2da Der	2da Der
Rango espectral (nm)	1638 – 1664 2118 – 2204	1624-1752 2100-2260	1832- 2490
Nº componentes PLS	2	5	6
SET CALIBRACIÓN			
Nº muestras	19	18	19
Laboratorio			
Concentración (mg/g)	12.03 – 35.99	0.443 – 1.219	0.135 – 0.267
RMSEC	0.389	0.032	0.008

RESULTADOS Y DISCUSIÓN

SET PREDICCIÓN				
Nº muestras	Laboratorio	14	8	10
	Producción	14	11	13
	Concentración (mg/g)	14.50 – 36.57	0.358 – 1.066	0.156 – 0.262
	RMSEP	0.569	0.062	0.035

Sin embargo, la determinación del contenido de Ketoprofeno y de los dos conservantes, en muestras expuesta al aire durante cierto tiempo, presenta errores crecientes con el tiempo de exposición al ambiente; este hecho puede asociarse a la evaporación que sufre el etanol durante la exposición al ambiente que provoca modificaciones en la composición química de la muestra. La predicción del API muestra valores superiores al valor nominal; sin embargo la predicción de los dos conservantes presenta valores por debajo del nominal. Para corregir esta fuente de error del modelo de cuantificación del Ketoprofeno, se prepararon 11 nuevas muestras mediante un diseño de experimentos (D-Optimal) de modo que la correlación espectral entre los analitos sea inferior a 0.75. Estas muestras fueron preparadas ampliando el rango de concentraciones de etanol para incorporar al modelo la variabilidad química del etanol. En la tabla 5.5 se muestran los parámetros para los 3 nuevos modelos de Ketoprofeno después de incorporar nuevas muestras al conjunto de Calibración.

Tabla 5.5. Parámetros de los nuevos modelos de calibración.

	KET	MET	PRO
Pretratamiento	2da Der	2da Der	2da Der
Rango Espectral (nm)	1638 – 1708	1620-1746	1832- 2490
	2118 – 2234	2100-2256	
Nº Factores PLS	3	6	6

RESULTADOS Y DISCUSIÓN

SET CALIBRACIÓN

Laboratorio	Nº muestras	29	22	25
Concentración (mg/g)		12.03 – 37.16	0 – 1.071	0 - 0.303
RMSEC		0.718	0.069	0.015

SET PREDICCIÓN

Laboratorio	Nº muestras	14	10	9
Producción		14	19	10
Concentración (mg/g)		14.50 – 36.47	0.560 – 1.019	0.129 – 0.279
RMSEP		0.797	0.833	0.211

Se ha ensayado la capacidad predictiva del nuevo modelo, determinando el contenido de Ketoprofeno en muestras en las que se ha evaporado parte del etanol presente dejando la misma muestra expuesta al ambiente durante un intervalo de 30 minutos y se registran 4 espectros cada 10 minutos. Se realizó un test ANOVA de una vía, siendo el parámetro a evaluar el tiempo al cual está la muestra expuesta al aire. Así, del primer modelo de Ketoprofeno muestra significación del tiempo de evaporación en los resultados mientras que el segundo modelo es insensible a la evaporación de parte de solvente en los resultados de cuantificación tanto del API como de los conservantes. De la misma manera se calcularon los valores de F para los dos modelos de cada uno de los conservantes, obteniendo el mismo resultado positivo.

Validación del método

El método NIR propuesto para la determinación simultánea del principio activo Ketoprofeno (KET), p-hidroxibenzoato de metilo (MET) y p-hidroxibenzoato de propilo (PRO) ha sido validado utilizando las guías ICH y EMEA, evaluando la selectividad, linealidad, exactitud, precisión (repetitividad y precisión intermedia) y robustez.

Selectividad

RESULTADOS Y DISCUSIÓN

La selectividad del método se establece mediante una identificación del gel previa a la cuantificación de los analitos. La identificación se realiza mediante la utilización de una biblioteca espectral que contiene 22 espectros del preparado procedentes de 7 lotes de producción: el criterio discriminante utilizado para identificación es el coeficiente de correlación de los espectros de segunda derivada en el intervalo espectral completo (1100 – 2498 nm). El *threshold* de 0.85 permite identificar correctamente el preparado farmacéutico y diferenciarlo del resto de productos.

La selectividad de la biblioteca ha sido evaluada identificando muestras externas a su construcción y todas las muestras de producción analizadas fueron identificadas positivamente. La rutina de análisis establecida implica que una muestra no identificada no puede ser cuantificada.

Tabla 5.6. Resultados obtenidos de la validación de los 3 modelos construidos.

Parámetro		KET	MET	PRO	
Linealidad	n	13	10	10	
	Concentración (mg/g)	14.50 – 36.47	0.560 – 1.219	0.156 – 0.262	
	Ordenada orígen	-1.38 ± 1.79	0.05 ± 0.211	-0.02 ± 0.03	
	Pendiente	1.06 ± 0.08	1.01 ± 0.22	0.93 ± 0.12	
	R	0.998	0.959	0.988	
Exactitud	n	28	29	23	
	Residuales (mg/g)	-0.23	0.015	0.002	
	DE	0.776	0.055	0.012	
	<i>t</i> experimental	1.127	1.472	0.986	
	<i>t</i> tabulada	2.056	2.048	2.074	
Repetitividad	Promedio NIR (mg/g)	25.40	0.722	0.190	
	CV (%)	1.89	8.01	8.56	
Precisión Intermedia	Día	<i>F</i> experimental	15.25	2.15	0.49
		<i>F</i> tabulada	19.00	19.00	19.00
	Analista	<i>F</i> experimental	0.55	5.35	0.36
		<i>F</i> tabulada	18.51	18.51	18.51
	CV (%)		1.23	3.32	6.03

RESULTADOS Y DISCUSIÓN

	n	28	31	29
Robustez	<i>t</i> experimental	1.076	1.754	1.608
	<i>t</i> tabulada	2.052	2.042	2.048

La validación del resto de propiedades se ha realizado siguiendo las guías indicadas anteriormente. La tabla 5.6 que recoge los datos relevantes de la validación. la validación de la robustez se ha realizado a partir del análisis los resultados obtenidos para muestras de más de 30 lotes de producción registradas durante un año; un test t de las diferencias entre los valores NIR y los de HPLC (método de referencia) indican que estos métodos no son significativamente distintos.

El análisis mediante espectroscopia NIR de muestras con una matriz muy absorbente de analitos que se encuentran en una baja concentración obliga a extremar la atención en la construcción de los modelos correspondientes. La alta correlación espectral entre los analitos ha sido corregida preparando muestras con la menor correlación entre las concentraciones y la reducida aportación al espectro de los componentes ha obligado a realizar una cuidadosa selección de las longitudes de onda y del modo espectral a utilizar para la construcción de los modelos. En este apartado se ha demostrado la capacidad de aplicar el NIR en el control industrial de este tipo de productos, presentando los modelos características de calidad apropiadas para el control analítico de estos productos.

Cuantificación de Dexketoprofeno en granulados farmacéuticos.

Uno de los procesos farmacéuticos más comunes es la granulación de la mezcla de las materias primas. Una vez granulado el producto se puede compactar para la formación de comprimidos. En el presente apartado se ha determinado el contenido de DexKetoprofeno en un granulado farmacéutico.

La elevada correlación espectral entre los componentes (ver tabla 5.7) hace imprescindible que la correlación entre las concentraciones de los componentes sea

RESULTADOS Y DISCUSIÓN

pequeña para obtener modelos robustos. Para conseguirlo se han preparado diferentes placebos, que cubren un rango de concentraciones del +/-10% de la concentración nominal, con una baja correlación entre los componentes; sin embargo, la suma de 2 excipientes (Sacarosa (Suc) y Aroma de Limón(Fla)) y el API constituyen un 99,8 % de la composición total del placebo (Glicirrinato Amónico (Gly), Neohesperidina (Neo) y Amarillo de Quinoleína (Qui), lo que hace ineludible la existencia de una alta correlación entre las concentraciones de estos dos excipientes.

Tabla 5.7. Coeficientes de correlación espectral entre todos los componentes de la preparación farmacéutica.

	API	Qui	Neo	Gly	Fla	Suc
API	1					
Qui	0.878	1				
Neo	0.936	0.950	1			
Gly	0.969	0.916	0.979	1		
Fla	0.937	0.919	0.990	0.979	1	
Suc	0.883	0.778	0.920	0.922	0.941	1

Un error en la dosificación de los excipientes durante la preparación industrial podría provocar errores importantes en la predicción del contenido de API. Esta situación debe evitarse para obtener modelos robustos, lo que se consigue mediante un diseño adecuado de la composición de las muestras que se prepararan en el laboratorio. Se prepararon 16 placebos distintos y después se tomaron aleatoriamente para mezclarlos con cantidades variables de API, obteniendo un total de 45 muestras que fueron distribuidas en los conjuntos de calibración y de validación de los modelos.

Estas muestras en polvo preparadas en el laboratorio cubren ampliamente el rango de variabilidad química que puede presentar el producto comercial y cada una de estas

RESULTADOS Y DISCUSIÓN

muestras fue granulada siguiendo el procedimiento descrito en la parte de metodología, obteniendo así un segundo conjunto de muestras que incorporan también la variabilidad física debida a la granulación (el tamaño de partícula obtenido es similar al del producto comercial).

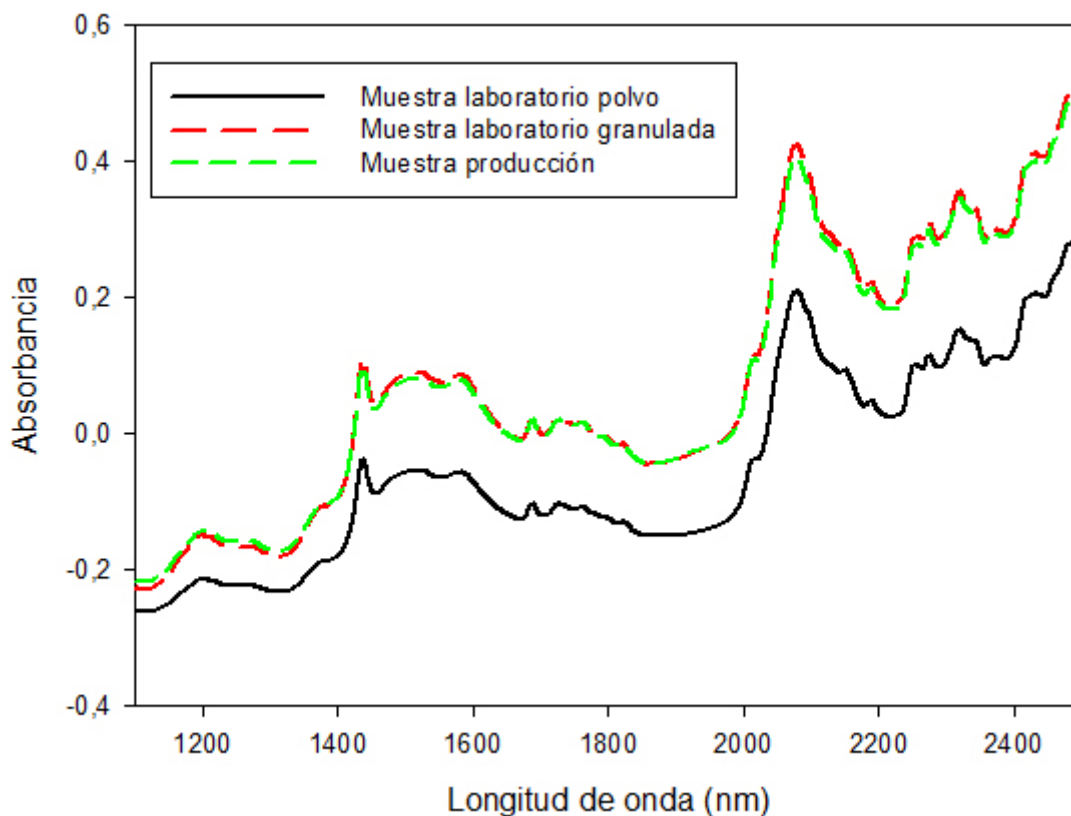


Figura 5.6. Espectro NIR de muestras de laboratorio, polvo y granulada, y de muestra de producción.

La Figura 5.6 muestra el espectro en modo Absorbancia de una muestra de producción, una de laboratorio en polvo con un contenido de API cercano al valor nominal, y la misma muestra granulada en el laboratorio. Se observa el desplazamiento entre ambos espectros causado por el diferente tamaño de partícula entre ambas muestras. Las pequeñas diferencias en el tamaño de partícula entre las dos muestras son consistentes con la similitud de los espectros

RESULTADOS Y DISCUSIÓN

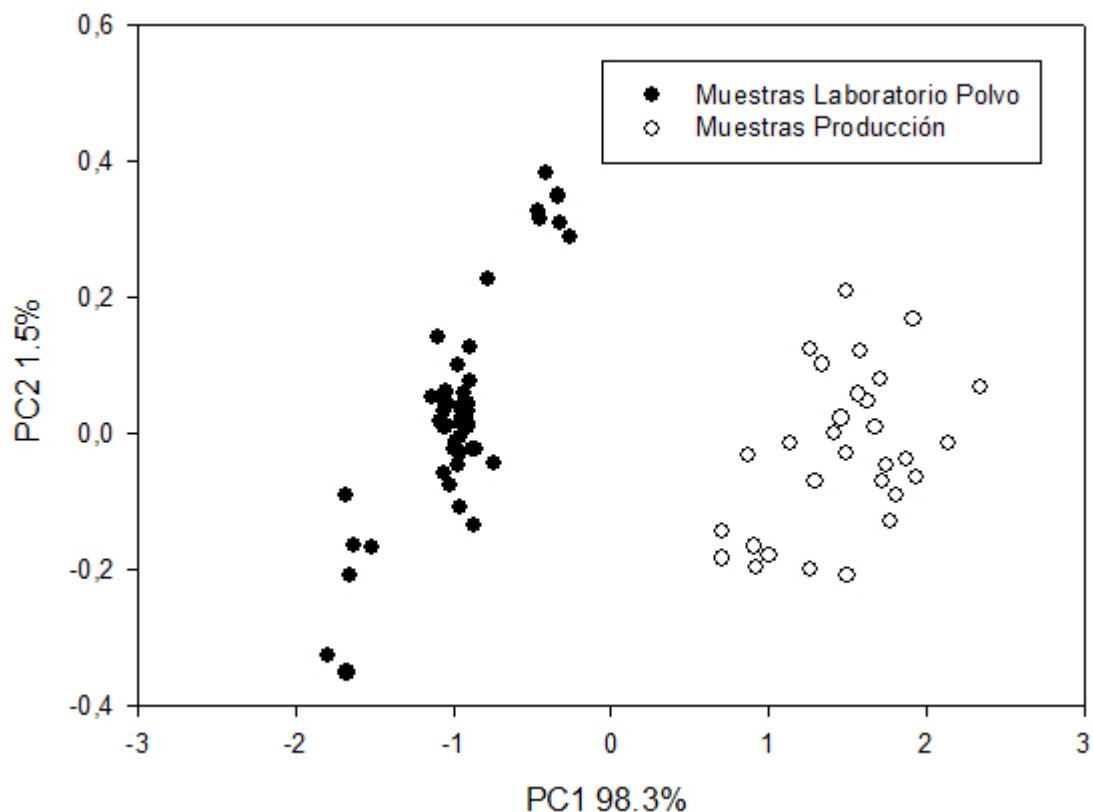


Figura 5.7. Gráfico de dispersión de scores del PCA de los espectros en Absorbancia de muestras de laboratorio en polvo y muestras de producción.

En primer lugar se planteó la construcción de un modelo de calibración para la determinación de API en muestras de producción, a partir de muestras en polvo preparadas en el laboratorio. El gráfico de dispersión de scores (Figura 5.7) de un PCA obtenido con espectros en modo Absorbancia de muestras de laboratorio en polvo y de producción muestra las diferencias originadas por el tamaño de partícula, las cuales ya se observaron en la figura 5.6. Las muestras se agrupan según su método de preparación en el PC1 con un 98.3 % de la varianza total, siendo la fuente de variabilidad más importante la diferencia en las propiedades físicas entre las muestras preparadas por pesada en el laboratorio y las muestras de producción. Durante el proceso de producción, el tamaño de partícula de las muestras aumenta, mientras que las muestras preparadas en el laboratorio se presentan en polvo, con un tamaño de partícula inferior, las muestras se distribuyen en el PC1 en función del tamaño de

RESULTADOS Y DISCUSIÓN

partícula. El PC2 (1.5% Varianza total explicada) recoge fundamentalmente las diferencias de concentración del API. Los gráficos de dispersión de los espectros después de diferentes pretratamientos espectrales (SNV y derivadas) muestran una dispersión semejante, lo que indica que los pretratamientos no eliminan el efecto del desplazamiento espectral debido a las diferencias en el tamaño de partícula, aunque si reducen su magnitud (no se muestran los gráficos).

En la Tabla 5.7 se muestran los parámetros del modelo y las predicciones obtenidas con muestras de laboratorio en polvo; el pretratamiento que mejores resultados aportó fue el de 2º Derivada de Savitzky-Golay (modelo 1). Se observa que con 5 factores PLS, se obtienen unos estadísticos de calibración correctos (98.5% varianza explicada) y una adecuada capacidad predictiva (RSEC=2.4%) para muestras de laboratorio en polvo. Sin embargo la predicción de muestras de laboratorio granuladas y muestras de producción no es correcta, obteniendo un error sistemático por defecto muy elevado, lo que puede atribuirse a las diferencias físicas (tamaño de partícula) entre ambos grupos de muestras.

La inclusión de muestras de producción en el conjunto de calibración pretende incorporar la variabilidad física y como consecuencia, mejorar la capacidad predictiva para muestras de laboratorio granuladas y de producción. El modelo 2 se ha construido añadiendo 7 muestras de producción al conjunto de calibración del modelo 1. Se han requerido dos factores PLS más que con el modelo 1, lo que se explica por la inclusión de muestras con diferencias espectrales, y aunque ha mejorado la predicción de las muestras de producción, no es todavía adecuada la capacidad predictiva del modelo (valores elevados del RSEP).

RESULTADOS Y DISCUSIÓN

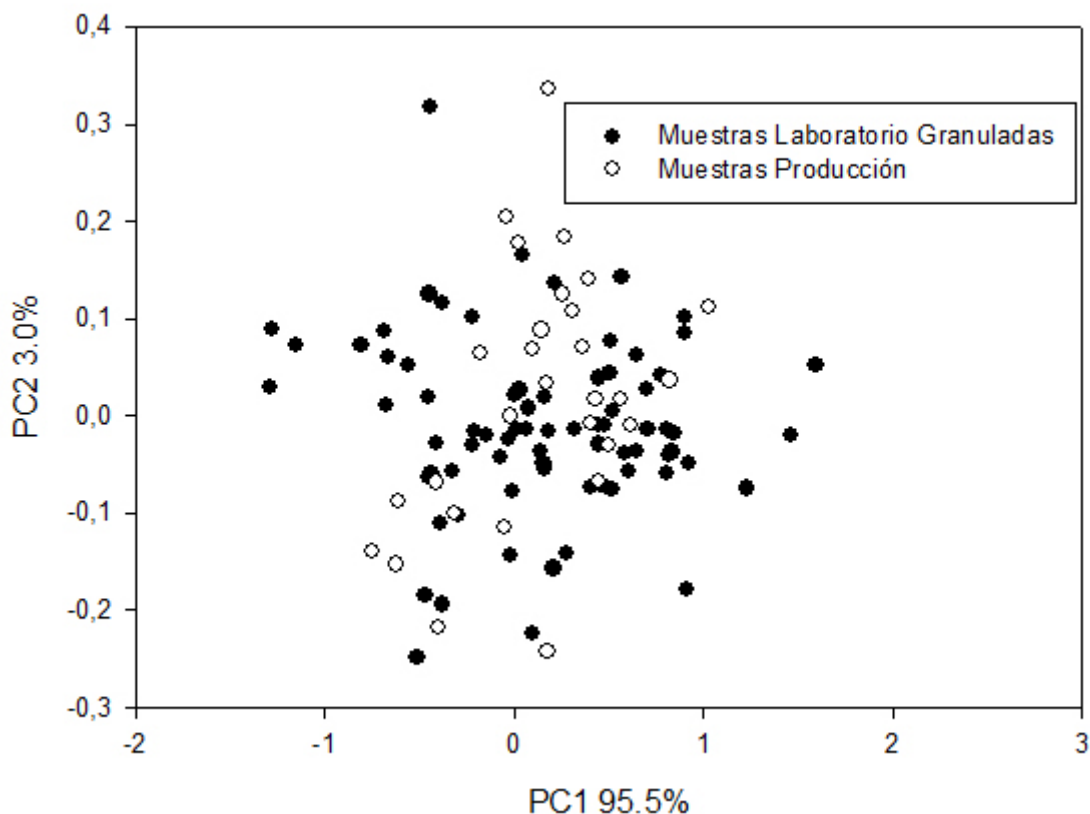


Figura 5.8. Gráfico de dispersión de los scores al realizar un PCA con los espectros de Absorbancia de muestras granuladas en el laboratorio y muestras de producción.

Con el objeto de mejorar los resultados obtenidos con los dos modelos anteriores se realizó un estudio similar con las muestras de laboratorio granuladas. El gráfico de dispersión de los dos primeros scores de los dos tipos de muestra (figura 6.3) indica que las diferencias espectrales debidas a parámetros físicos entre ambos tipos de muestras han prácticamente desaparecido. Las muestras se han agrupado en un único conjunto, lo que indica que ambos tipos de muestras presentan las mismas fuentes de variabilidad espectral. El PC1 (95,5%) está asociado a la concentración de API. Las muestras de producción están agrupadas en la parte central del cluster como consecuencia del estrecho rango de concentración de API que presentan (+5%). Sin embargo las muestras de laboratorio granuladas presentan un amplio rango de concentración de API que produce una mayor distribución de los scores a lo largo del eje del PC1. Este gráfico confirma la correcta preparación de las muestras de laboratorio que recoge toda la variabilidad de las muestras de producción.

RESULTADOS Y DISCUSIÓN

Tabla 5.7. Valores estadísticos de los modelos PLS construidos en función del tipo de muestras utilizadas.

Modelo	Muestras	Pretratamiento	Region (nm)	Calibration							Predicción				
				Factores			Laboratorio				Producción				
				PLS	Var Y%	RSEC%	Polvo		Granulado		Residuales				t exp
							RSEP%	n	RSEP%	n	RSEP%	n	Promedio	S.D.	
1	Laboratorio	2nd Derivative	1100-1200	5	98.5	2.4	5.6	18	66.5	45	44.8	30	-3.97	1.81	12.01*
	Polvo		1600-2500												
2	Laboratorio	2nd Derivative	1100-2500	7	96.7	3.2	8.6	18	46.2	45	8.5	23	0.23	0.82	1.51**
	Polvo & Producción														
3	Laboratorio	2nd Derivative	1100-1200	5	98.7	2.0	--		3.9	20	1.6	30	-0.05	0.13	2.00*
	granuladas		1400-2500												

* t crit = 2.04 (95%, n=30)

** t crit = 2.07 (95%, n=23)

RESULTADOS Y DISCUSIÓN

Se construyó un modelo de calibración (modelo 3) exclusivamente con muestras de laboratorio granuladas. En la Tabla 5.7 se muestra como los estadísticos de calibración son correctos (5 factores PLS, RSEC=2.0%) y su capacidad predictiva, RSEP= 3,9 y 1,6 % para las muestras de laboratorio granuladas y de producción respectivamente. Este modelo combina la sencillez en cuanto a factores PLS del modelo 1 y aún mejora la capacidad predictiva del modelo 2.

La Figura 5.9 muestra un gráfico de residuales donde se observa la evolución de la exactitud con muestras de producción, de los modelos en función de las muestras empleadas para construir dichos modelos. El modelo 1 muestra unos residuales sistemáticos muy elevados. Este modelo no incluye la variabilidad física del proceso de producción y clara muestra de ello es este error sistemático tan elevado. Los residuales obtenidos con el modelo 2 son aleatorios alrededor de cero, sin embargo en algunos casos son cercanos al 20%. Finalmente, el modelo 3 presenta valores de residuales muy cercanos a cero y con una distribución aleatoria.

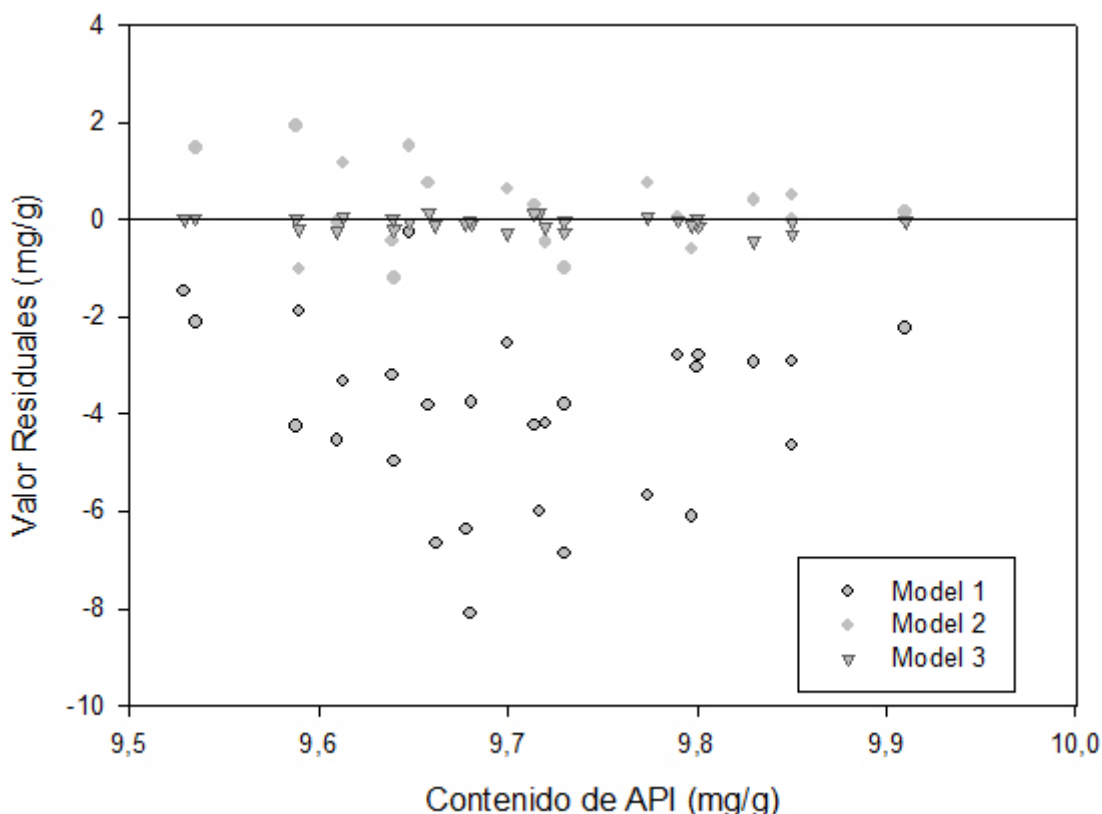


Figura 5.9. Gráfico de residuales de muestras de producción.

RESULTADOS Y DISCUSIÓN

La nueva estrategia de preparación de las muestras de calibración, que incorpora en este conjunto la variabilidad propia del proceso de producción, permite obtener modelos más simples y con una adecuada capacidad predictiva, haciendo innecesaria la inclusión de muestras de producción. Además, los valores de referencia del contenido de API en las muestras se han establecido a partir de los pesos de los componentes de las mezclas preparadas en el laboratorio, haciendo innecesario la aplicación del método de referencia para su determinación, que es la metodología habitual en los modelos NIR.

DETERMINACIÓN DE TRANSFORMACIÓN POLIMÓRFICA EN UN GRANULADO FARMACÉUTICO

En una reciente publicación [40] se ha demostrado la transformación polimórfica a una forma amorfa que experimenta el Dexketoprofeno durante el proceso de granulación húmeda en presencia de celulosa microcristalina. La semejanza del proceso de granulación aplicado en este caso nos ha llevado a realizar un estudio para establecer si también ahora se producía esta transformación.

Para este estudio de posible transformación polimórfica del API se prepararon 7 mezclas binarias de Dexketoprofeno, que se encuentra en el formato Polimorfo B) y Sacarosa (a diferencia del trabajo citado donde se utilizó celulosa microcristalina) en un rango de concentraciones de API de 0 a 20 mg·g⁻¹. Alícuotas de cada muestra fueron sometidas al mismo proceso de granulación descrito anteriormente y se registraron sus espectros NIR antes y después de ser granuladas.

Se ha aplicado el algoritmo MCR-ALS a las muestras binarias de API y sacarosa en polvo y granuladas siguiendo la misma metodología que en el estudio anterior con la finalidad de observar una posible transformación polimórfica del API durante el proceso de granulación.

Se escogió todo el rango de longitudes de onda utilizando los modos espectrales de absorbancia, SNV y 1^a y 2^a Derivada y se aplicaron las restricciones de no-negatividad

RESULTADOS Y DISCUSIÓN

de concentración y de closure. Además se aportó como información adicional la concentración de sacarosa en cada muestra y su espectro. La descomposición se aplicó tanto a las muestras en polvo como a las muestras granuladas. Los perfiles espectrales del API obtenidos de las muestras en polvo y granuladas se han comparado con los espectros de API polimorfo B y el asignado a la forma amorfa del mismo.

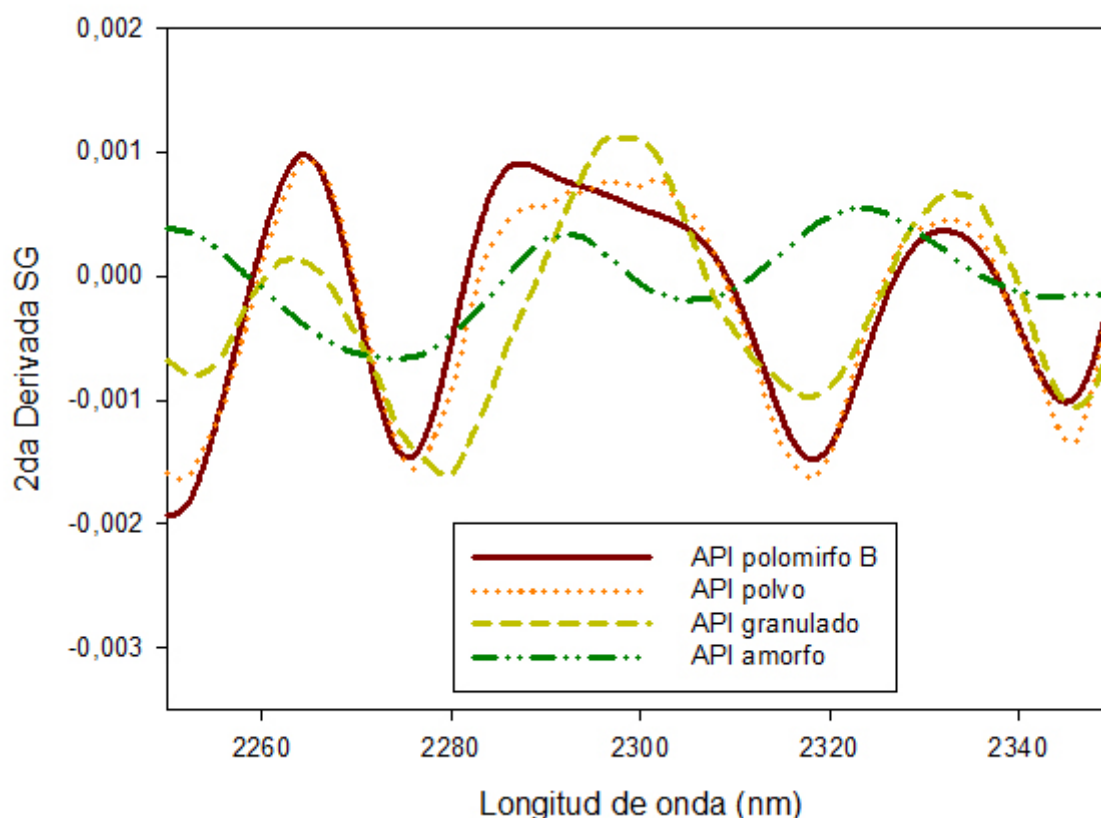


Figura 5.10. Espectros NIR en 2da Derivada del API (forma amorfa y polimorfo B), y perfiles espectrales de API en muestra en polvo y granuladas obtenidos en la descomposición MCR-ALS.

Los perfiles espectrales obtenidos por MCR-ALS de las muestras en polvo y granuladas son muy similares al API puro, aunque el de muestras granuladas muestra un desplazamiento debido al proceso de granulación. Las máximas diferencias quedan resaltadas en la región de 2240-2360 nm (figura 5.10). La similitud entre los perfiles espectrales y el espectro de API polimorfo B se demuestra por la semejanza de los coeficientes de correlación entre los espectros en el intervalo de longitudes de onda de 2240-2360 nm. El valor obtenido entre el espectro de API polimorfo y el API

RESULTADOS Y DISCUSIÓN

granulado es 0.91, mientras que el coeficiente de correlación entre el API amorfo y el API granulado es 0.28. La correlación con el espectro de la forma amorfa es mucho menor, lo que indica que no ha tenido lugar transformación polimórfica durante la granulación.

La aplicación del algoritmo MCR-ALS a mezclas granuladas de Dexketoprofeno y sacarosa demuestra que no tiene lugar la transformación polimórfica del primero a diferencia de lo que ocurre cuando la granulación se realiza con celulosa; este estudio permite establecer que las características del soporte sobre el que se realiza la granulación son las que determinan la transformación.

5.2 DETERMINACIÓN PARÁMETROS DURANTE PROCESO (*ON LINE MONITORING*)

En el estudio que se muestra a continuación se han monitorizado algunos procesos, de los muchos que se realizan semanalmente en la planta de producción, sin modificar las condiciones normales de trabajo y sin extraer muestras durante la evolución del mismo (con la excepción de los procesos que se utilizaron para construir modelos cuantitativos). Esta monitorización del proceso nos permite obtener información en tiempo real del estado del mismo reduciendo y/o eliminando fuentes de variación o inconvenientes: granulación inadecuada, secado excesivo, etc. que, sin duda, tendrían un efecto importante en la calidad del producto final. Por tanto, desarrollar una buena metodología de trabajo en la que se consiga realizar el control de proceso a tiempo real se presenta con un gran atractivo para el progreso de las metodologías PAT en la industria farmacéutica.

En primer lugar se evaluó el procedimiento de registro de espectros y la necesidad e importancia de la fuente de iluminación externa en la calidad de los espectros obtenidos. Varios procesos fueron registrados situando la fuente de luz y la sonda de fibra óptica en diferentes posiciones, y con ángulos distintos. Una única lámpara halógena proporciona suficiente iluminación para obtener espectros reproducibles y sin ruido espectral. Además, colocando la sonda y la luz en un ángulo aproximado de 45° , y la sonda tocando la ventana del granulador, permite registrar espectros nítidos y con una correcta repetitividad. La figura 5.11 muestra una fotografía del granulador de lecho fluido con el montaje del espectrofotómetro. Alrededor de la ventana circular del granulador, se coloca un marco cuadrado metálico que tiene como función principal la de soporte de la sonda de fibra óptica y de la luz halógena. Esta disposición se desmonta después del registro de cada proceso y se coloca de nuevo en cada uno de los procesos siguientes.

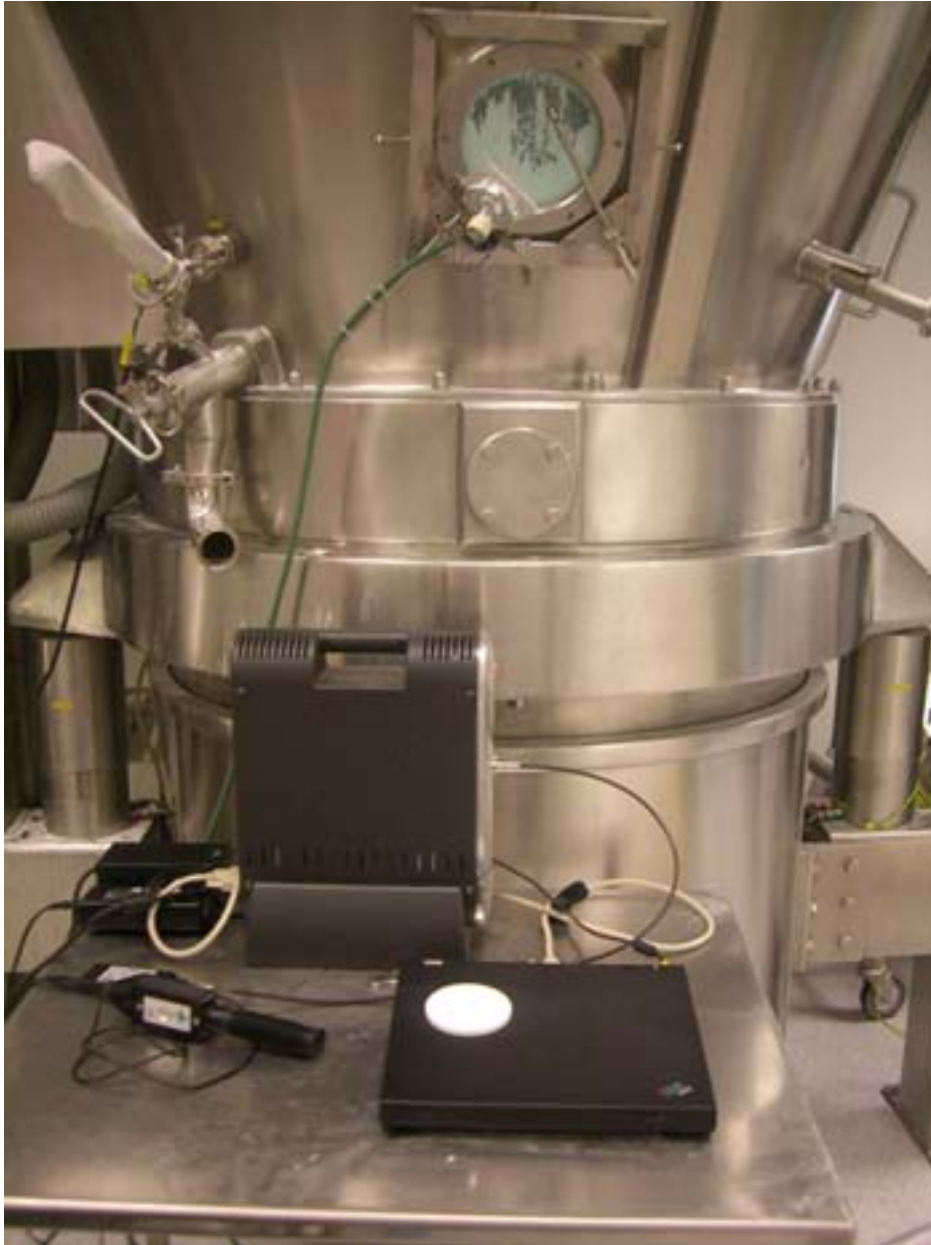


Figura 5.11. Imagen del montaje utilizado para realizar la monitorización del proceso de granulación. La sonda de fibra óptica está colocada en la ventana del reactor, junto con una fuente de luz halógena.

El estudio de un proceso de granulación se ha estructurado en función del tipo de información obtenida: en primer lugar se realizó un estudio de la evolución de los espectros registrados durante los procesos. Este estudio cualitativo se ha realizado mediante un Análisis en Componentes Principales (PCA) de todos los espectros.

RESULTADOS Y DISCUSIÓN

El segundo estudio tiene como finalidad realizar un análisis cuantitativo destinado a obtener modelos de calibración apropiados para predecir correctamente los parámetros de interés, humedad, tamaño de partícula y densidad a lo largo de todo el proceso de granulación para determinar, cuantitativamente, el estado del proceso en cualquier momento construyendo modelos PLS adecuados para la predicción de dichos parámetros.

5.2.1 INFORMACIÓN CUALITATIVA

Los resultados que se presentan corresponden al estudio de 5 procesos. La figura 5.12 muestra todos los espectros recogidos durante un proceso de granulación. 1 espectro correspondiente a cada una de las 3 etapas del proceso donde se observa la variación en las bandas de absorción durante las etapas del proceso de granulación, entre las que destacan el aumento de absorción de las bandas en las regiones de alrededor de 1450 y de 1950 nm debido a la adición de ligante durante la etapa de granulación y su posterior disminución durante el secado del granulado. La importante señal de absorción en la región de las longitudes de onda cortas, entre 1000 y 1300 nm, es debida principalmente a la absorción del vidrio de la ventana.

RESULTADOS Y DISCUSIÓN

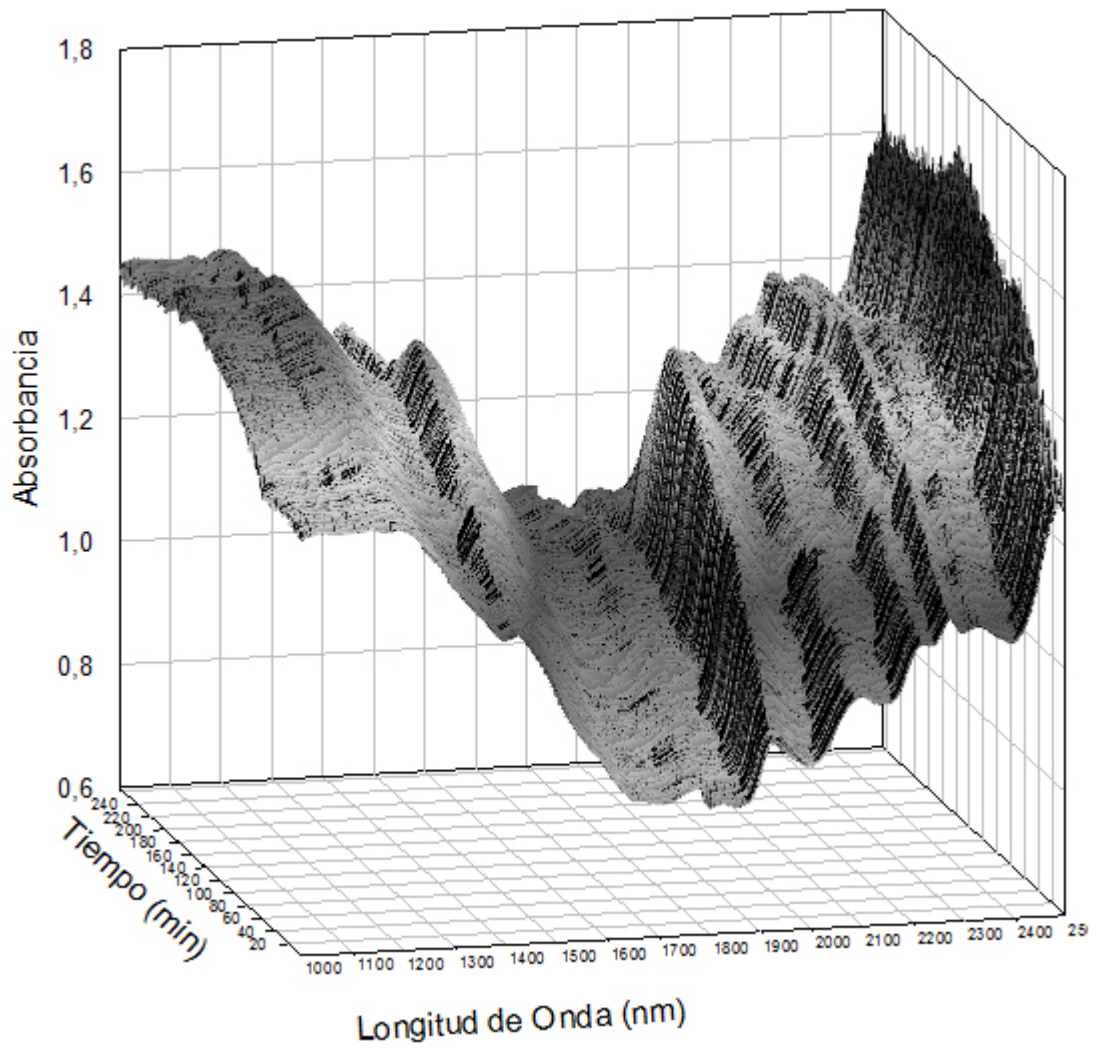


Figura 5.12 Espectros NIR de un proceso de granulación

Un PCA de la matriz de todos los espectros de un proceso en modo Absorbancia muestra que dos primeros factores PC recogen el 99% de la varianza de todo el proceso.

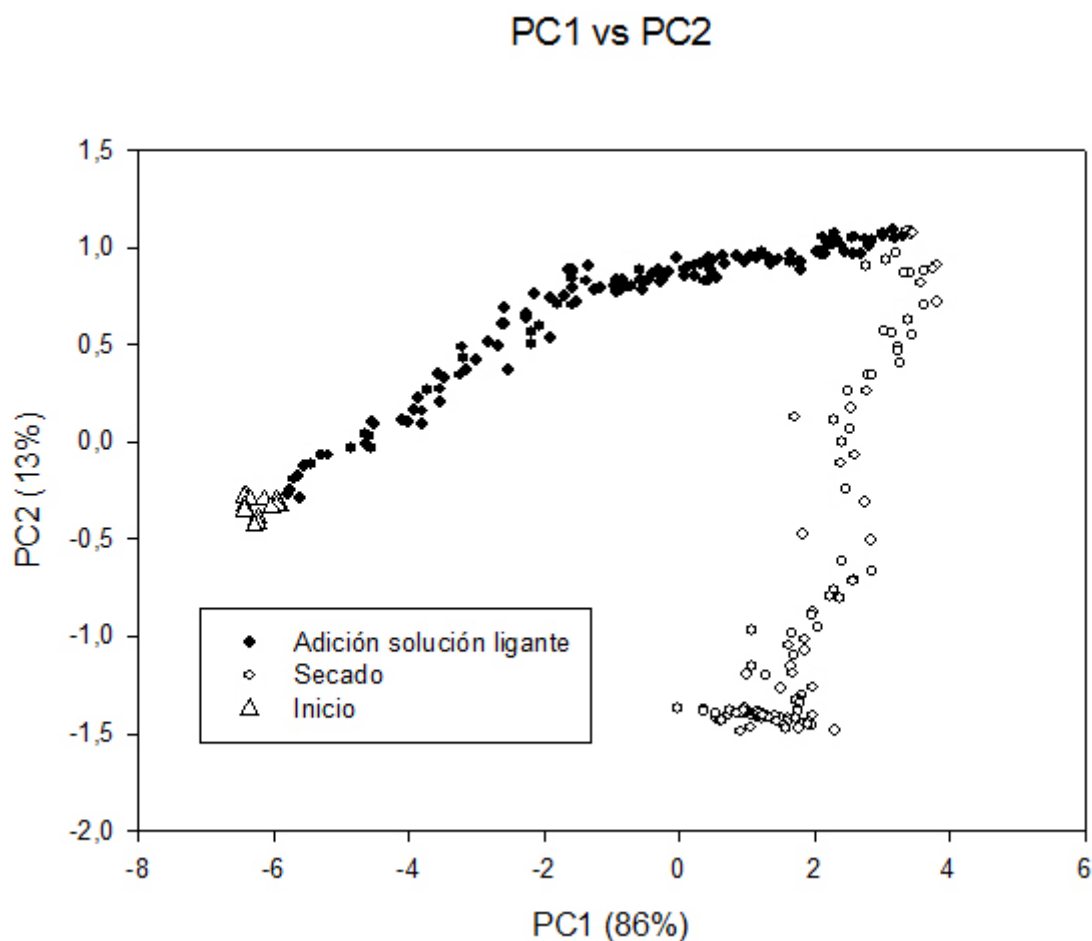


Figura 5.13. Gráfico de dispersión de una PCA de todos los espectros de un proceso en modo Absorbancia. Los scores están representados en función de la etapa correspondiente.

En la figura 5.13 se han representado el gráfico de dispersión de los scores PC1 vs PC2 de un proceso en la que se han representado en función de la etapa del proceso. En la parte inferior izquierda de la figura 5.13, aparecen los scores que corresponden a la etapa inicial de agitación, que están agrupados sin mostrar ninguna evolución. A partir del minuto 15, en el que empieza la adición de solución ligante se produce una evolución de los scores tanto en el PC1 como en el PC2 hasta aproximadamente el minuto 70 en que solo se observa la evolución del valor del score del PC1. Finalmente, durante la etapa de secado hay una evolución importante del score del PC2 aunque también se aprecia una leve variación en el PC1. En este gráfico se distinguen claramente las tres etapas del proceso de granulación.

RESULTADOS Y DISCUSIÓN

Todos los procesos registrados presentan perfiles de evolución de los espectros y de los scores muy similares al mostrado en la figura 5.13, y por tanto este gráfico permitiría visualizar si un proceso se está realizando bajo las condiciones normales de operación (NOC) y el proceso de granulación se está realizando de manera correcta.

Sin embargo, al desarrollarse todos los procesos en las condiciones habituales de trabajo no se ha contemplado la posibilidad de modificarlas hasta situaciones de fuera de control y no ha sido posible observar la utilidad del PCA como herramienta para el control global del proceso relacionando la evolución de los scores al estar algún parámetro fuera de control.

El PC1 representa un 86% de la variabilidad total por lo que el parámetro que esté relacionado a este factor será el de mayor contribución al proceso. Un gráfico interesante es el de los valores de los scores en función del tiempo a que se ha registrado el espectro.

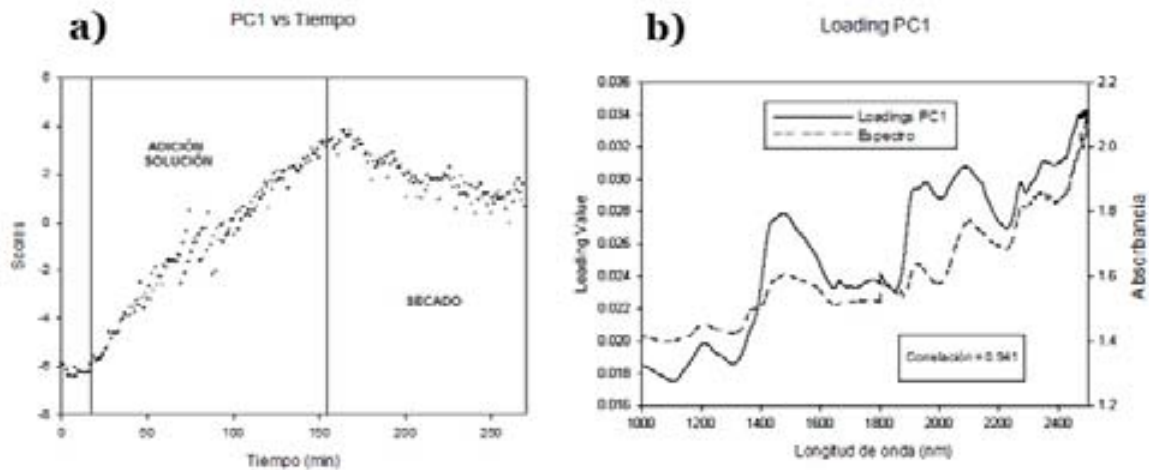


Figura 5.14. a) Evolución de los scores del PC1 en función del tiempo. b) Representación del espectro de la muestra y los loadings espectrales del PC1.

En la figura 5.14 se han representado los scores del PC1 frente al tiempo del proceso señalando las tres etapas del proceso: agitación de la muestra y calentamiento, adición de solución ligante y secado del granulado.

RESULTADOS Y DISCUSIÓN

Una interpretación de la figura atendiendo a la evolución de los scores con el tiempo nos indica que el valor de los scores se mantiene constante durante la primera etapa donde tiene lugar únicamente un proceso de mezcla de los excipientes y principio activo, por lo que no se espera evolución en los scores. El aumento de los scores durante la segunda etapa (adición de solución ligante) puede explicarse por el proceso de granulación; la adición de la solución provoca la formación de agregados en la mezcla. Esta variación del tamaño de partícula provoca un desplazamiento del espectro que se ve reflejado en la evolución de los scores. Este desplazamiento ya se había observado en la figura 6.12. Durante la tercera etapa hay un aumento de la temperatura del flujo de aire sin adición de solución ligando. En esta etapa se produce una disminución del valor del score que es coherente con la disminución del tamaño de partícula producido por la colisión entre ellas y la consiguiente reducción del tamaño medio.

El loading espectral del PC1 (Figura 5.14 b) presenta un perfil altamente parecido al espectro de la muestra obtenido; el coeficiente de correlación entre el loading y el espectro de la muestra es superior a 0.94. Este hecho, junto con la interpretación de los scores, hace pensar que el primer factor está asociado al cambio de tamaño de partícula debido al proceso de granulación.

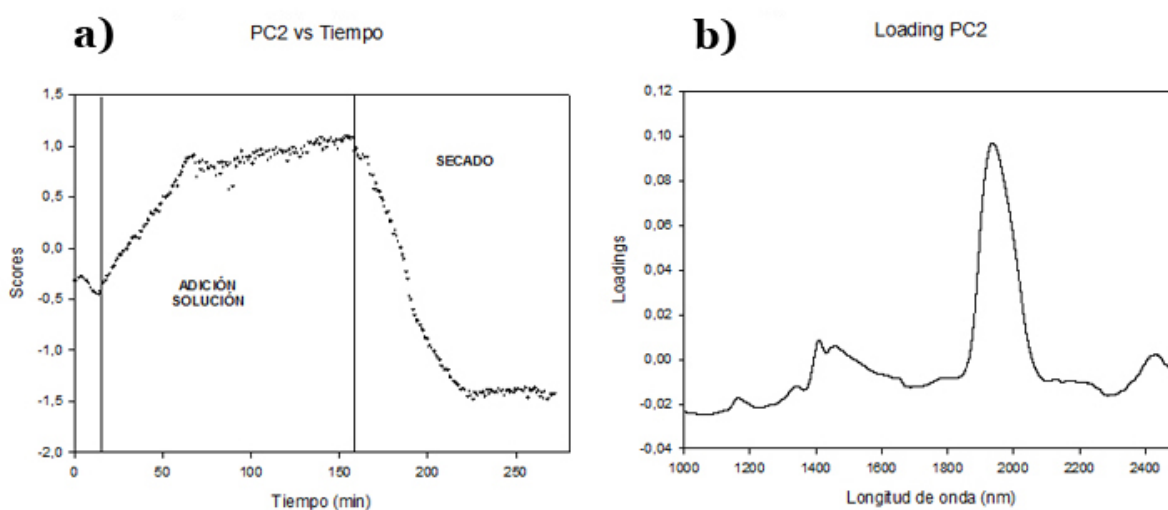


Figura 5.15. a) Evolución de los scores del PC2 en función del tiempo b) Loadings espectrales del PC2

RESULTADOS Y DISCUSIÓN

Un análisis de la evolución del PC2 permite extraer también conclusiones interesantes: la figura 5.15a muestra la evolución del valor de los scores del PC2 aumentando en la segunda etapa debido a la adición de solución acuosa ligante y que disminuye en la tercera etapa a consecuencia del secado del granulado. La menor pendiente del gráfico a partir del minuto 70 indica una menor incorporación de agua al sistema ya que una parte se pierde inmediatamente evaporada por el flujo de aire caliente. Incluso puede deducirse que el agua residual es menor que la de los componentes al inicio del proceso (menor valor del score). También el loading del PC2 (figura 5.15b) muestra una gran semejanza con el espectro del agua, lo que induce a pensar que el segundo factor está relacionado con el contenido de agua de la muestra en el proceso.

Prosiguiendo con el estudio de los score del PCA, se estudiaron los PC 3 y 4, que recogen cerca de un 1 % de la variabilidad explicada; sin embargo la evolución de los scores con el tiempo, así como los loadings no aportaron información relevante.

Por tanto, la evolución de los scores en el PC1 y PC2 puede ser utilizada como parámetro general para conocer la evolución de los procesos y establecer un control de los mismos. El estudio descrito se puede extender al resto de procesos seguidos ya que en todos ellos se observa el mismo comportamiento; la única diferencia es el valor de los scores pero los gráficos de evolución de los mismos son idénticos.

5.2.2 INFORMACIÓN CUANTITATIVA

Los parámetros monitorizados en el proceso de producción para este fármaco son: la distribución del tamaño de partícula, el contenido en agua y la densidad. Estos tres son los parámetros de interés para el control de este proceso definidos por la empresa. Sin embargo, otros parámetros podrían haber sido también monitorizados.

Determinación distribución tamaño de partícula

Desde el punto de vista de la fabricación del producto final se considera imprescindible un 80% de la totalidad del lote de producción con un tamaño de partícula comprendido

RESULTADOS Y DISCUSIÓN

entre 125 y 1000 micras. Sin embargo, además de construir un modelo para este intervalo, también se han construido modelos para tamaño de partícula inferior a 125 y superior a 1000 micras para saber cuanta muestra presenta un tamaño de partícula indeseable y que pueda producir segregación en el caso de tamaño de partícula inferior a 125 micras, y obturaciones en el caso de más de 1000 micras, en la etapa de compactación.

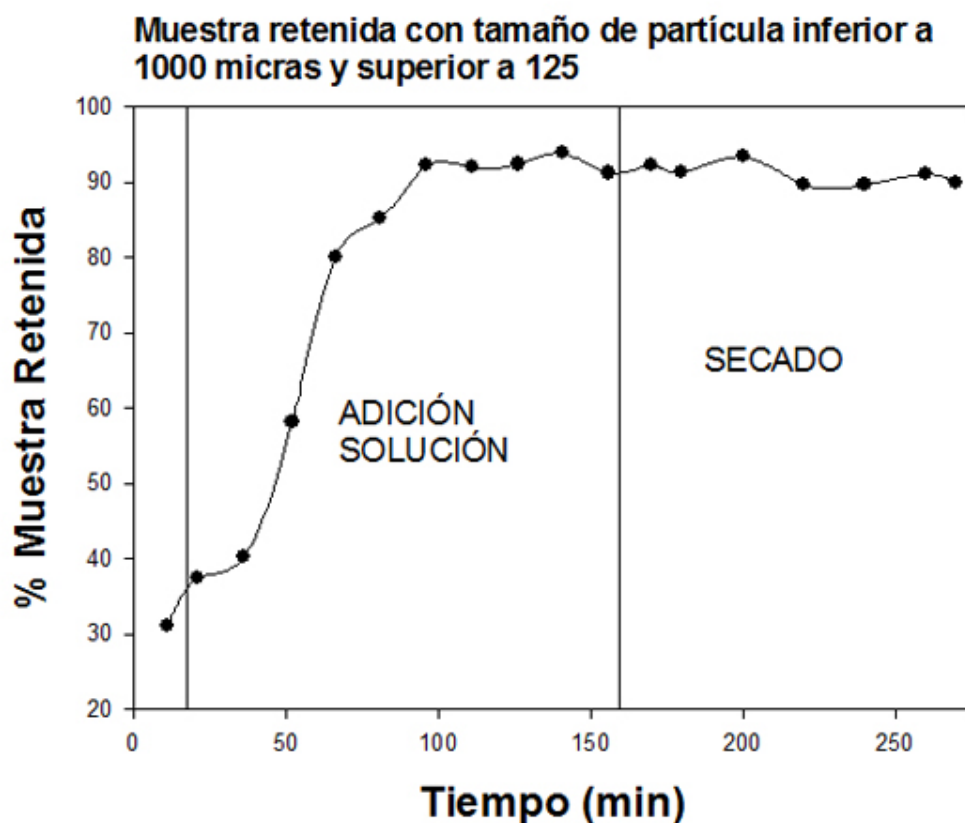


Figura 5.16. Evolución de los valores de referencia del contenido en agua y del porcentaje de muestra con tamaño de partícula inferior a 1000 micras.

En la figura 5.16 se muestra la evolución del porcentaje de muestra con un tamaño de partícula inferior a 1000 micras (límite superior) y superior a 125, obtenidos por método de referencia a las muestras extraídas. Este gráfico muestra una gran similitud con la evolución de los scores de las figura 5.14. Al principio del proceso, gran parte de la muestra presenta un tamaño de partícula inferior a 125 micras, correspondiente al tamaño de partícula de las materias primas. Cuando empieza la adición de la solución

RESULTADOS Y DISCUSIÓN

ligante se van formando gránulos grandes tal como se ve en la figura. En esta etapa un 5% de muestras supera las 1000 micras de tamaño de partícula. De aquí la final de la adición el tamaño de partícula medio aumenta. Una vez terminada la adición de solución ligante empieza la etapa de secado en la que la agitación provoca colisiones entre los gránulos que causan una disminución del tamaño de partícula y aumenta la fracción de muestra retenido en los tamices de malla más pequeña. Estos cambios no se observan prácticamente en el gráfico debido a que recoge un intervalo de tamaño de partícula amplio. Sin embargo, comparando los porcentajes en cada uno de los tamices utilizados (los valores no se muestran), el porcentaje en los tamices de 710 y 500 micras disminuye durante la etapa de secado, aumentando los porcentajes de 250 y 180 micras.

Se han construido modelos PLS para determinar la evolución del tamaño de partícula durante el proceso de granulación húmeda. Se han agrupado en 3 conjuntos: % retenido superior a 1000 micras, % retenido entre 1000 y 125 micras, y finalmente el resto, inferior a 125 micras. Por debajo del tamiz de malla inferior (38 micras) hay menos de un 0,1 % de muestra. Se escogieron espectros de 3 procesos para preparar el modelo de calibración, utilizando espectros de los otros procesos para validar el modelo. Se construyeron un total de 3 modelos PLS1 de calibración para los tres intervalos de partícula, cuyos parámetros más significativos se muestran en la tabla 5.8. Todos los modelos se construyeron en modo Absorbancia y sin realizar ningún pretratamiento espectral.

La figura 5.17 (a, b y c) muestra la predicción de un proceso no implicado en la calibración con los 3 modelos PLS construidos y superpuestos los valores del parámetro obtenido de cada una de las muestras extraídas del proceso. A pesar de las diferencias en los primeros 50 minutos, debidas a la poca renovación de muestra en la ventana del granulador, a partir de este momento, puede observarse un buen ajuste entre ambos valores lo que demuestra la buena capacidad predictiva de los tres modelos construidos.

RESULTADOS Y DISCUSIÓN

Tabla 5.8. Parámetros relevantes de los modelos de calibración contruidos para la determinación de tamaño de partícula, contenido en agua y densidad aparente.

PARÁMETROS DE ESTUDIO					
	Tamaño de partícula			% Agua	Densidad Aparente
	>1000 μm	1000-125 μm	< 125 μm		
Pretratamiento Espectral		Absorbancia		SNV	Absorbancia
Rango Espectral (nm)	1000-1400 1500-1875 2100-2500	1000-1215 1630-1900 2150-2425	1000-2500	1800-2000	1485-2375
Factores PLS	5	4	4	3	3
RMSEC	0.6 %	2.8 %	3.7 %	0.5 %	0.01 mg/L
RMSEP	1.9 %	11.8 %	12.0 %	1.2 %	0.03 mg/L

RESULTADOS Y DISCUSIÓN

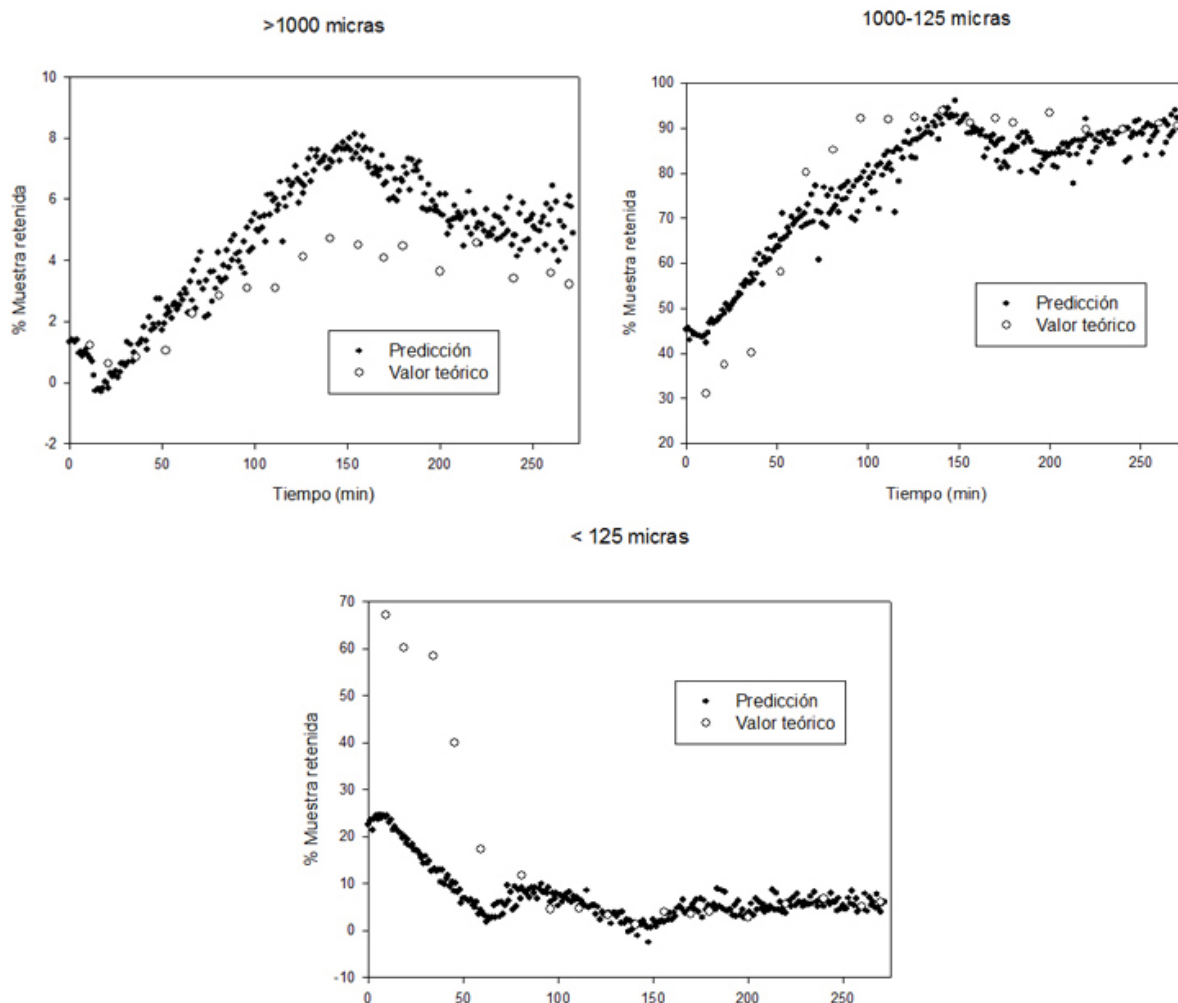


Figura 5.17 Predicción del % de muestra en los 3 intervalos de tamaño de partícula modelados para un proceso externo a la calibración.

Aproximadamente a partir del minuto 180 se alcanza el 80 % de muestra con un tamaño entre 125 y 1000 micras. A partir de este punto prácticamente no se observa evolución de los valores en ninguno de los 3 intervalos de tamaños de partícula, lo que indica que el proceso de granulación ha concluido. Esta estabilidad se presenta también en los Scores del PC1, y puede ser identificado como un factor indicador de que el proceso ha finalizado. Es importante que al final del proceso los porcentajes de los dos intervalos extremos sean pequeños y que sea máximo el porcentaje de granulado en el intervalo de 125 a 1000 micras.

RESULTADOS Y DISCUSIÓN

El control del tamaño de partícula en este tipo de procesos se presenta como una etapa esencial para que el granulado pueda pasar a la siguiente etapa de compresión, ya que variaciones en las variables como son el caudal de aire o de adición de la solución ligante, la temperatura, etc. pueden influir en la obtención del tamaño de partícula deseado.

Seguimiento de la evolución de la humedad

En el estudio cualitativo se demostró la asociación entre el PC2 y el contenido de humedad de la muestra en granulación y se decidió estudiar la construcción de un modelo PLS1 para cuantificar el contenido de humedad durante el proceso.

Como ya ocurriese con el tamaño de partícula, hay un valor establecido frontera. En este caso el producto final no puede contener más de un 5% de contenido en agua al finalizar el proceso de granulación.

Para reducir el efecto de la variación del tamaño de partícula en los espectros se aplicó el pretratamiento SNV a la matriz de espectros, y a semejanza del estudio anterior se utilizaron los espectros de 3 procesos para la calibración y los otros 2 procesos restantes para la validación externa.

RESULTADOS Y DISCUSIÓN

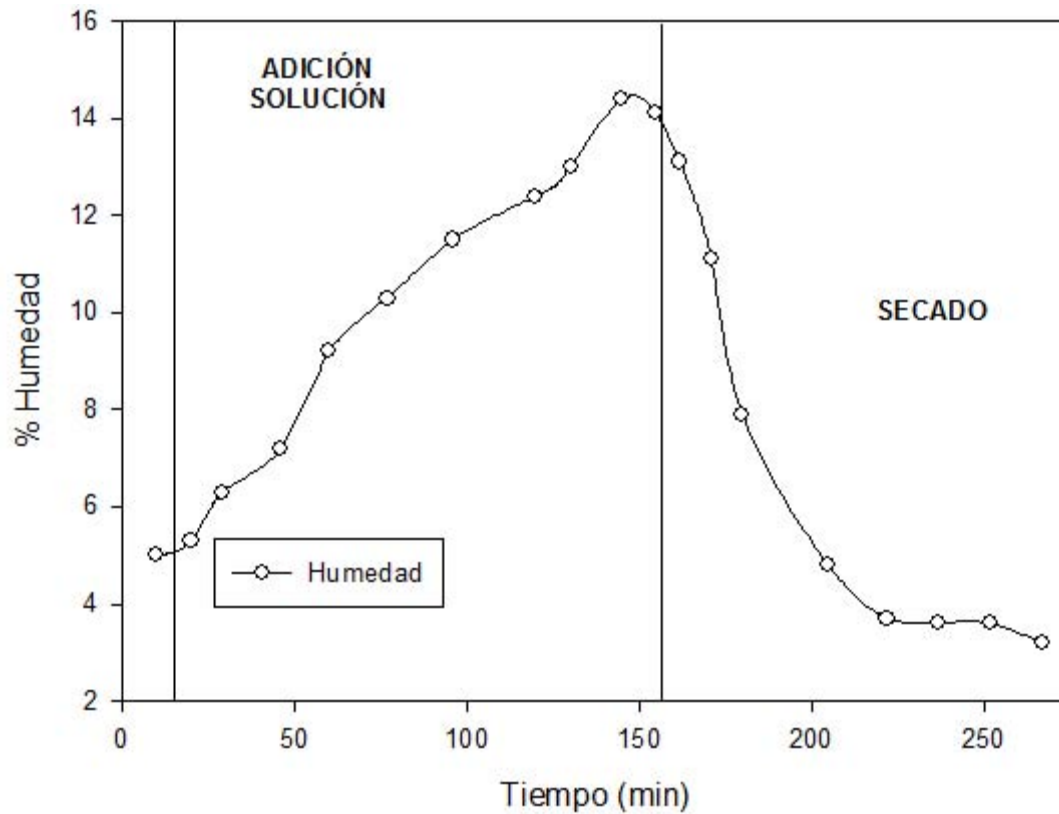


Figura 5.18. Evolución de los valores de referencia del contenido de agua en un proceso.

La tabla 5.8 muestra los parámetros más significativos del modelo PLS construido. Aunque el primer factor PLS describe un 95% de la varianza total, son necesarios un total de 3 factores (99%) para ajustar correctamente el modelo y conseguir una buena capacidad predictiva.

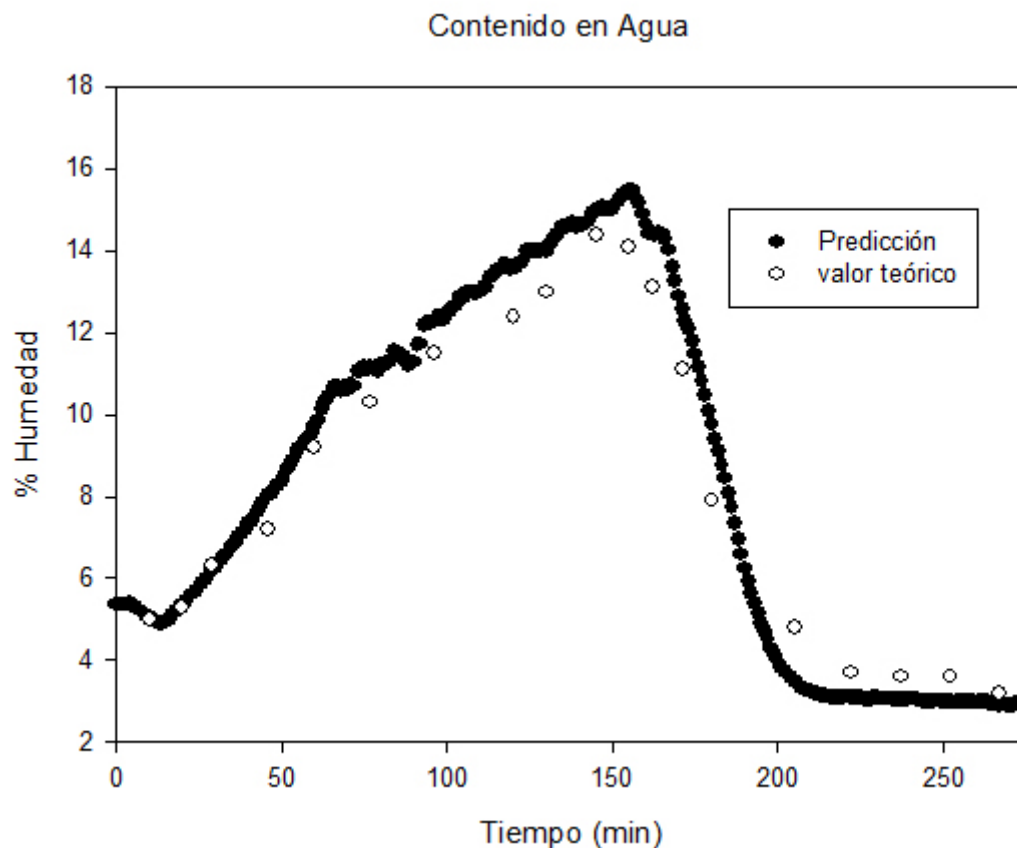


Figura 5.19 Predicción del contenido de agua en un proceso externo a la calibración.

La figura 5.19 muestra la predicción de la evolución del contenido en humedad calculados, con el modelo desarrollado y aplicado a un proceso, que no ha sido incluido en la calibración y también los valores de referencia; el granulado alcanza el valor del 5% de humedad en el minuto 200 aproximadamente, estabilizándose hasta el 3% durante el tiempo restante del proceso.

Seguimiento de la densidad.

El último parámetro monitorizado ha sido la variación de la densidad aparente de la muestra durante el proceso. De la misma manera que en los anteriores apartados el valor de densidad aparente fue determinado en las muestras extraídas durante los procesos.

RESULTADOS Y DISCUSIÓN

Se ha construido un modelo de calibración con muestras de 2 procesos, y ha sido evaluado con otros procesos no incluidos en el modelo. Los parámetros más relevantes se muestran en la tabla 5.8.

En la figura 5.20 se muestra la predicción de un proceso externo a la calibración junto con sus valores de referencia. La evolución de la densidad en el producto sigue las tendencias, a veces contrapuestas, de la evolución del tamaño de partícula y del contenido en humedad de las muestras. Los valores obtenidos en la determinación realizada en las muestras extraídas siguen la misma tendencia que los valores predichos con el modelo construido. La estabilización del valor de la densidad aparente, a partir del minuto 220 aproximadamente, coincide además con la estabilización de los parámetros de tamaño de partícula y contenido de agua.

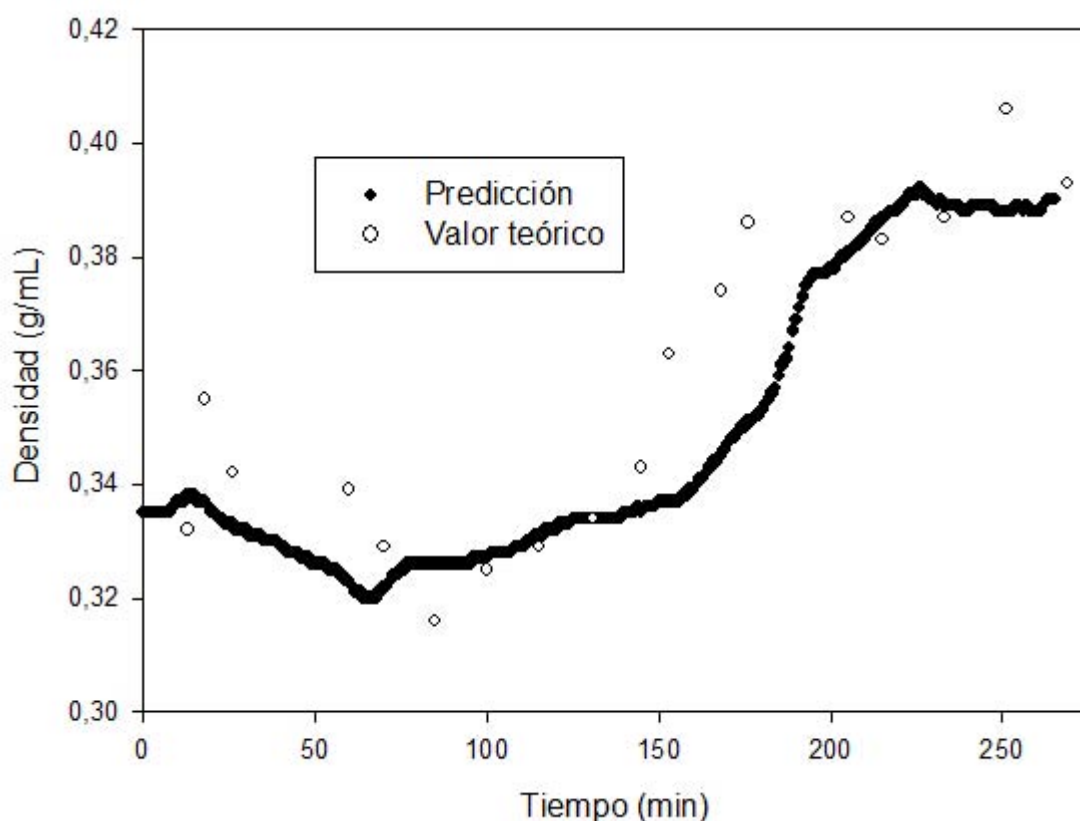


Figura 5.20. Predicción de la densidad aparente en un proceso externo a la calibración.

Los límites de control están entre 0,35 y 0,45, por lo que, tal y como se muestra en la figura, el proceso se encuentra dentro de las condiciones óptimas al final del mismo.

RESULTADOS Y DISCUSIÓN

La densidad aparente esta directamente relacionada con los parámetros de humedad y de tamaño de partícula, por lo que variaciones en el proceso quedan reflejadas en el seguimiento de los 3 parámetros monitorizados.

5.3 ESTUDIOS INICIALES NIR IMAGEN

La elevada cantidad de información que proporcionan las imágenes hiperespectrales (1 espectro cada píxel) permite obtener más de 90000 espectros por cada imagen. Por ello es necesario conocer que herramientas permiten tratar tal cantidad de datos para extraer la información relevante. Conocer las ventajas y desventajas del uso de las herramientas quimiométricas utilizadas dentro del campo de imagen NIR según la bibliografía será clave para hincar una nueva vía de investigación en el grupo.

Disponer de muestras cuya composición y distribución es perfectamente conocida permitirá saber si la información obtenida se corresponde con la información real.

Se ha dividido el último trabajo descrito en función de la información obtenida, por lo que se ha realizado un análisis de tipo cualitativo y otro cuantitativo.

Todos los resultados mostrados a continuación han sido aplicando SNV y suavizado de Savitzky-Golay con ventana de 11 puntos, una vez optimizados los pretratamientos que permiten obtener mayor información.

5.3.1 ANÁLISIS CUALITATIVO

Dentro del siguiente apartado se ha aplicado una de las herramientas más utilizadas en el estudio preliminar. Además se ha realizado un estudio de clusters utilizando 2 algoritmos diferentes con el objetivo de localizar y clasificar los analitos en la imagen. Finalmente, la utilización de una herramienta sencilla como es el coeficiente de correlación muestra la contribución de cada analito en la imagen.

Análisis PCA

RESULTADOS Y DISCUSIÓN

Los resultados de aplicar un PCA en las 3 muestras heterogéneas se muestra en la figura 5.21. En los 3 casos, los 2 primeros componentes principales explican más de un 99,8%.

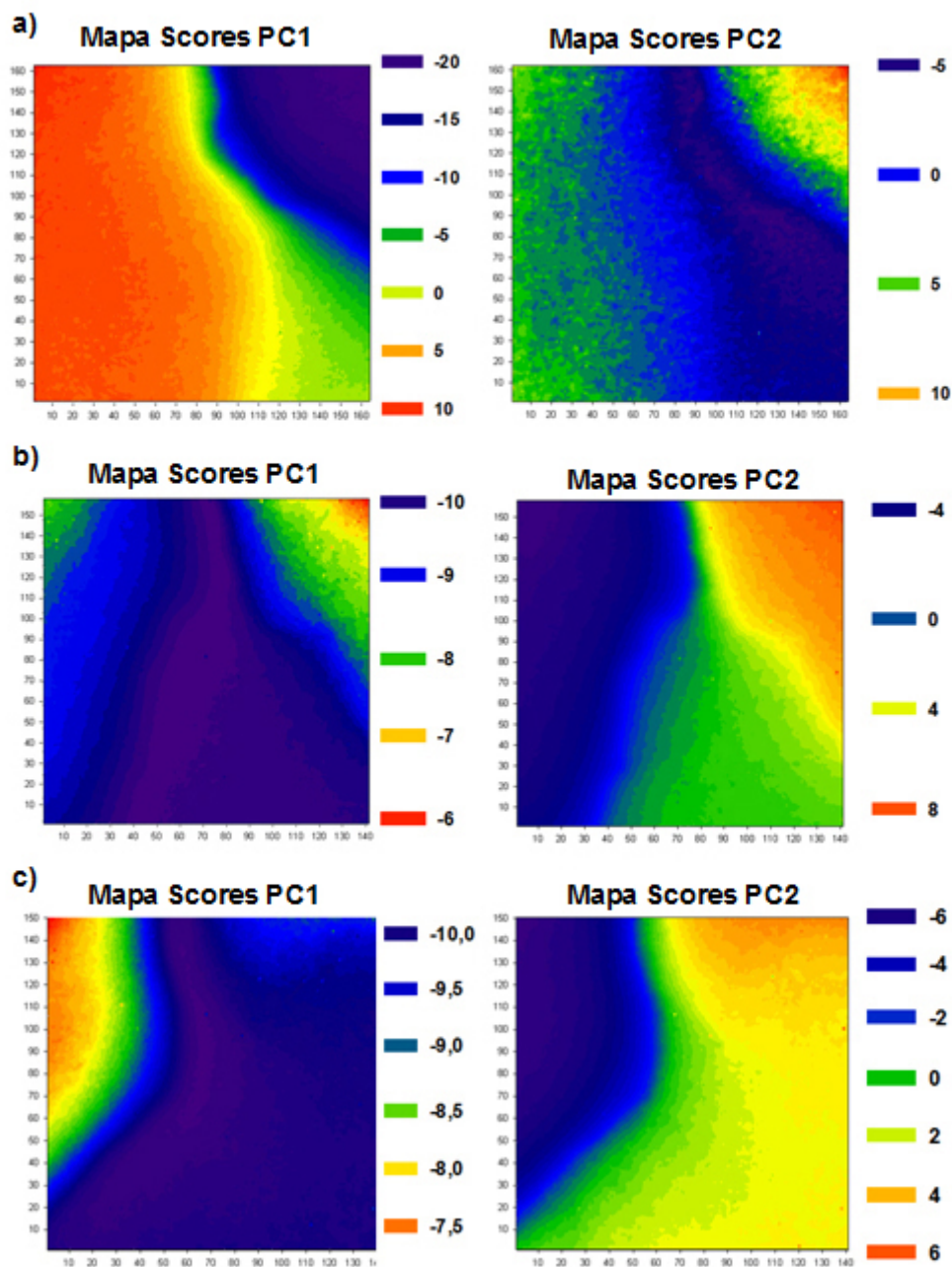


Figura 5.21. Mapas de scores de PC1 y PC2 para muestras heterogéneas.

Se podría esperar la posibilidad de poder relacionar 2 componentes principales con las 2 regiones con analito puro, además de un tercer componente principal asociado a la región de mezcla de analitos. Sin embargo, en los 3 casos, los loadings del tercer PC

RESULTADOS Y DISCUSIÓN

presentan un perfil muy ruidoso y el porcentaje de varianza explicada es inferior al 0,01%, por lo que no puede asociarse a información química relevante.

Analizando las 3 muestras individualmente se puede observar, en la muestra 1 (75% ibuprofeno y 25% almidón en la región de mezcla), los loadings espectrales del primer PC están altamente correlacionados con el espectro puro de ibuprofeno (tabla 5.9).

Tabla 5.9 Tabla de correlaciones entre loadings espectrales obtenidos en el PCA y los espectros puros de los 2 analitos de interés. En negrita se han marcado el coeficiente de correlación más alto para cada Loading.

Sample	Loading	Ibuprofeno	Almidón
1	1	0.944	0.559
	2	-0.313	0.818
	3	0.015	0.015
2	1	-0.795	-0.788
	2	-0.595	0.601
	3	0.015	0.020
3	1	-0.635	-0.908
	2	-0.763	0.402
	3	0.068	0.027
4	1	-0.888	-0.671
	2	-0.192	0.381
	3	0.002	0.097

Además, el mapa de scores del primer PC permite visualizar correctamente las 2 regiones de analito puro y la de mezcla. Los loadings espectrales del segundo PC están altamente correlacionados con el espectro puro del almidón, analito minoritario; además, en el mapa de scores de este PC, se muestra con más intensidad (color rojo) en la región con presencia más importante de almidón.

Para la muestra 2 (50% ibuprofeno y 50% almidón), los valores de correlación entre los loadings del primer PC y los espectros puros de ibuprofeno y almidón son similares

RESULTADOS Y DISCUSIÓN

pero con signo negativo. Además, el mapa de scores muestra una mayor intensidad en la región correspondiente al almidón. Esto es debido a las ambigüedades rotacionales que pueden ocurrir al desdoblar la matriz tridimensional, aplicar un PCA, y volver a redoblar la matriz. De esta manera hay que interpretarlo desde el punto de vista que la región con más relevancia en el PC1 es la región de mezcla de analitos. El segundo PC muestra un mapa de scores donde la zona más intensa corresponde a la región del almidón. Además, el coeficiente de correlación espectral es mayor para este analito.

Para la última muestra heterogénea (25% ibuprofeno y 75% almidón), cabría esperar que el primer PC mostrase la región del almidón con mayor intensidad. Sin embargo, esto no ocurre e incluso es la región del ibuprofeno la que aparece con valores más elevados de scores. Al comparar los loadings espectrales del primer PC con los espectros puros observamos una mayor correlación con el almidón, aunque ambos con valores negativos. Al analizar el segundo PC, algo parecido al primer PC ocurre. El mapa de scores muestra valores más intensos para la región de almidón y la de mezcla, con mayor proporción de almidón, mientras que los loadings espectrales esta inversamente correlacionados con el ibuprofeno. De nuevo nos encontramos con ambigüedades rotacionales, que necesitan de una correcta interpretación por parte del analista.

Este es uno de los mayores inconvenientes de utilizar el PCA en técnicas de imagen, pudiendo llevar a una incorrecta interpretación al analista al visualizar los mapas de scores.

RESULTADOS Y DISCUSIÓN

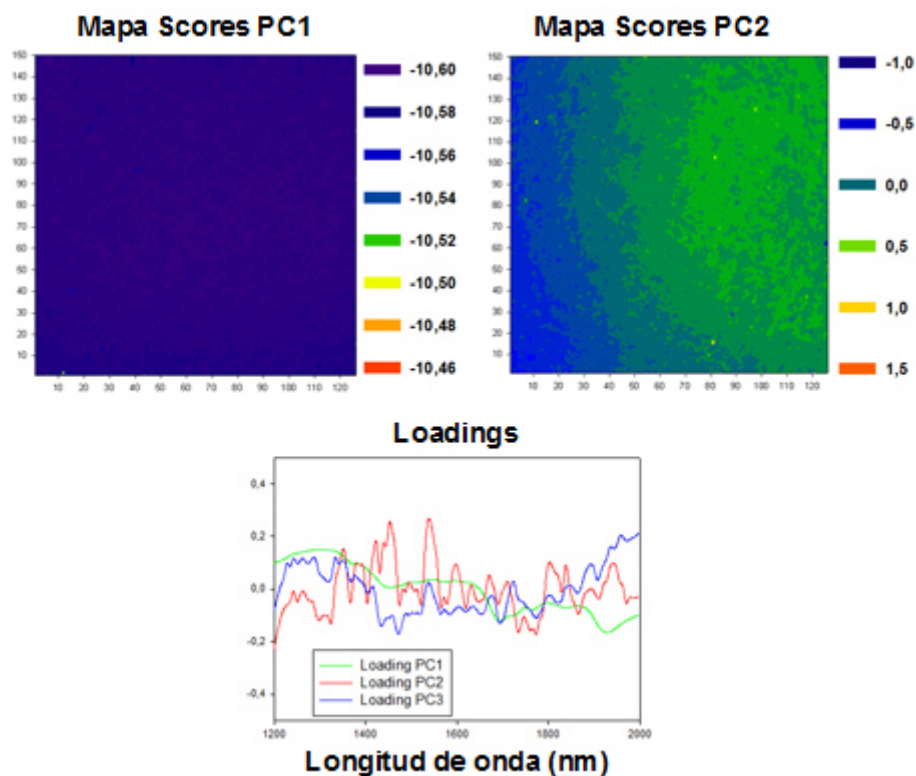


Figura 5.22. Mapas de scores y loadings espectrales de PC1 y PC2 para la muestra heterogénea.

La figura 5.22 muestra los resultados de aplicar un PCA en la muestra homogénea. El primer PC representa un 99,96 % de la varianza total, mientras que el segundo PC únicamente representa un 0,02%. Este hecho puede ser indicativo de tener una mezcla correctamente mezclada donde el PCA es incapaz de poder distinguir mapas exclusivos para ambos analitos. El gráfico del primer loading (en verde) se asemeja a una combinación lineal de los dos analitos. Los coeficientes de correlación entre los 2 espectros puros y los loadings se muestran de nuevo en la tabla 5.8, obteniendo una mayor correlación con el ibuprofeno, lógico debido a la mayor presencia de este analito en la muestra (65%). El mapa de scores del PC2 muestra una imagen prácticamente dividida por la mitad, con una región en azul y otra en verde. De los loadings espectrales no podemos relacionarlo con ningún componente químico, por lo que puede suponerse, previa aplicación de otras herramientas, a una fuente de variación física.

RESULTADOS Y DISCUSIÓN

Con todos estos resultados, aún conociendo la composición y distribución de las 3 muestras binarias, las ambigüedades rotacionales pueden guiar a una incorrecta interpretación de los datos. En el caso de la muestra homogénea, que se asemeja a las muestras que se pueden encontrar en casos reales en la industria farmacéutico, el PCA no aporta información significativa. Además, la elevada correlación espectral entre los 2 analitos dificulta la utilización del PCA.

Análisis de Segmentación. Clusters

Los clusters han sido ampliamente utilizados en la clasificación de muestras. Debido al elevado número de espectros contenidos en una imagen, la aplicación de clusters en imágenes podría separar segmentos dentro de la imagen en función de la semejanza espectral

Se ha calculado, para cada muestra, el número de clusters óptimo mediante el valor de Mean Silhouette y se han representado tal y como se muestra en la figura 5.23 siendo el valor óptimo de clusters el que contiene el máximo valor de Mean Silhouette.

Para las muestras 1 y 3, donde la región de mezcla presenta elevadas concentraciones de cada analito en cada una de las muestras, el número óptimo de clusters es 2. En la muestra 1, el área de la mezcla está asociada al cluster del ibuprofeno puro (figura 5.23), mientras que para la muestra 3, la región de mezcla está asociada a la región del almidón. Para la muestra 2, donde la región de mezcla contiene la misma proporción de los dos analitos, el número óptimo es 3, un cluster para cada una de las regiones debido a que la región de mezcla presenta diferencias significativas respecto a las regiones de los analitos puros, asignándole un cluster específico para esa región.

RESULTADOS Y DISCUSIÓN

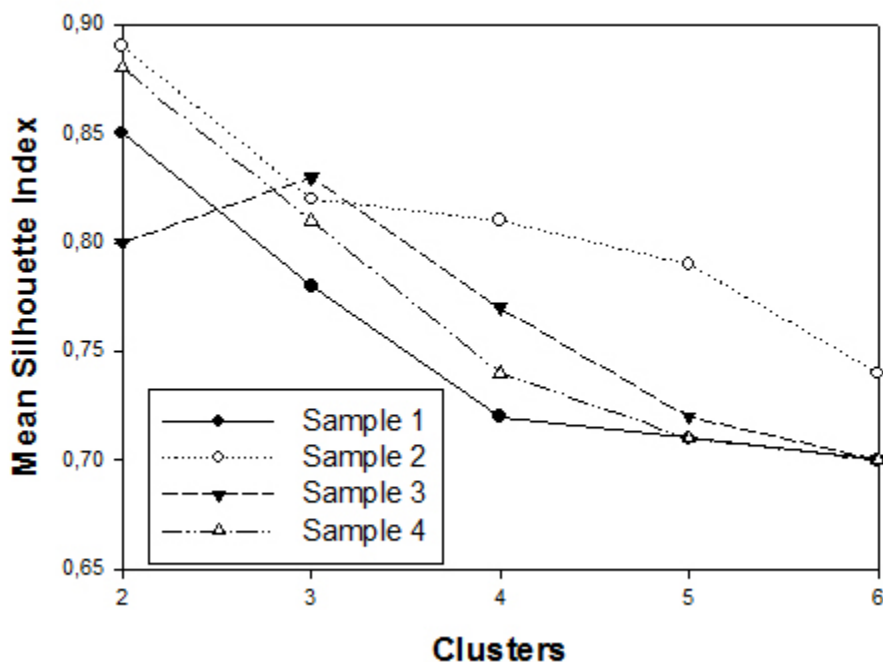


Figura 5.23. Valores del Silhouette Index en función del número de clusters par alas 4 muestras.

Los centros de los clusters (no se muestran) para las muestras 1 y 3 presentan un coeficiente de correlación con los espectros de los analitos puros muy elevado 0.989. Las regiones con mezcla estan asociadas al cluster que corresponde al analito con mayor presencia en la mezcla. Esto significa que el análisis de clusters no permite diferenciar espectralmente, para estas muestras la región de mezcla con la región del analito puro y mayoritario en la región de mezcla.

RESULTADOS Y DISCUSIÓN

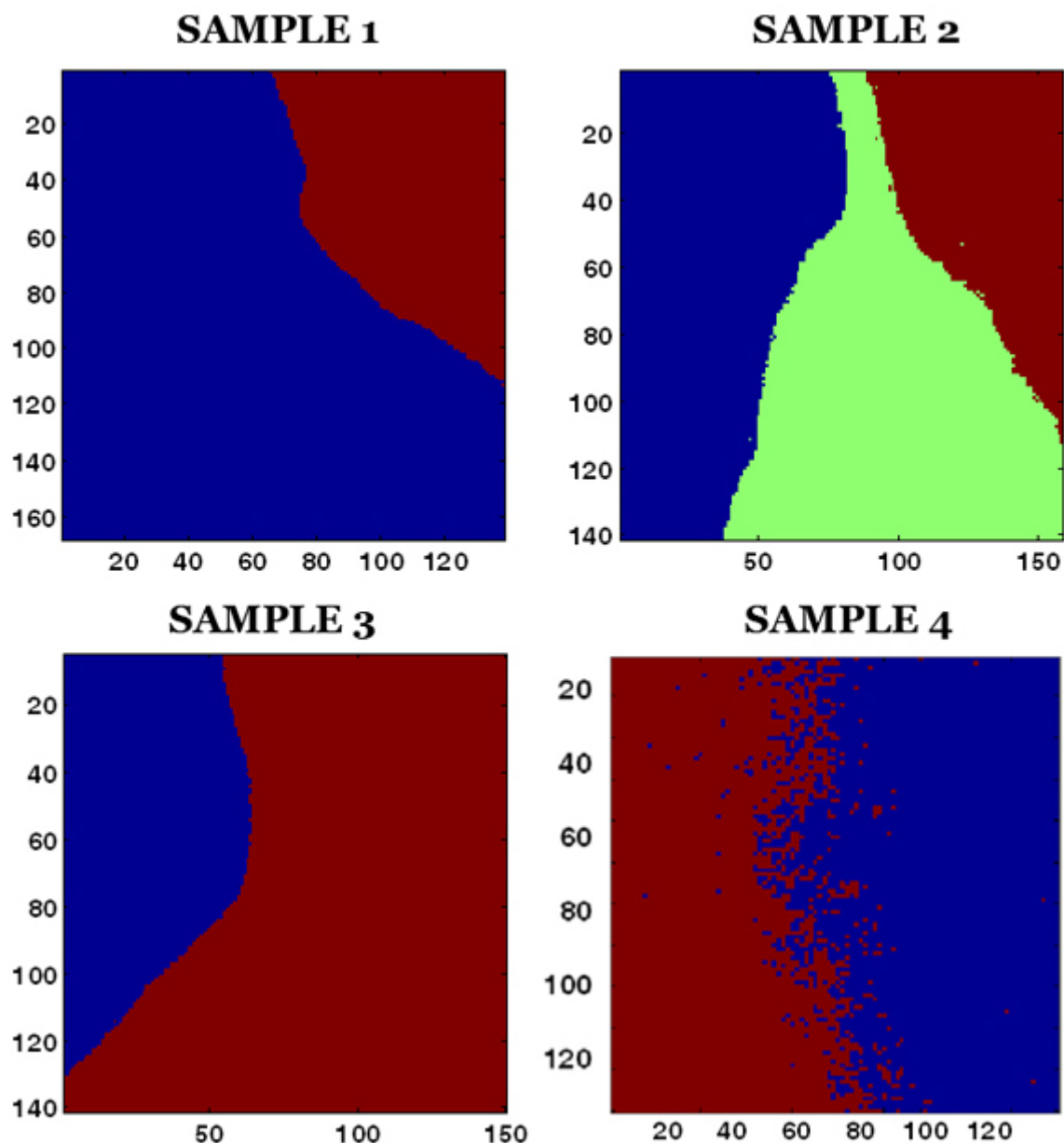


Figura 5.24. Representación de los clusters óptimos para las 4 muestras del estudio.

Los centros obtenidos en los clusters para la muestra 3 permiten explicar mucho mejor la distribución de los analitos en la superficie, diferenciando la zona de mezcla de las regiones de analitos puros. Esta diferenciación es debida a que la concentración en la zona de mezcla es de 50% cada analito, obteniendo diferencias espectrales significativas con las regiones de analito puro. En las muestras 1 y 3, aún conociendo la existencia de 3 regiones de diferente composición, podríamos haber aplicado 3 clusters. Son embargo, como ya se apreció en la figura 5.23, el número óptimo de clusters es 2, y or lo

RESULTADOS Y DISCUSIÓN

tanto, sin conocer la naturaleza de las muestras, los resultados obtenidos con esta clasificación hubiesen sido erróneos.

Del análisis de la muestra 4, que ha sido bien mezclada, cabe esperar un solo cluster, y del cual no se obtiene información. En el mejor de los casos, en función de la resolución espacial podríamos esperar un mapa con puntos de dos colores (en este caso azul y rojo por defecto) aleatoriamente distribuidos. Sin embargo, cuando se aplican 2 clusters aparece el mapa de la figura 5.24, donde la imagen está prácticamente dividida por la mitad. Comparando los centros de estos dos clusters se obtiene casi el mismo coeficiente de correlación con ambos analitos, por lo que los dos clusters no presentan diferencias químicas. Esta distribución, es parecida a la obtenida en el mapa de scores del segundo PC en la figura 2, y podría asociarse a algún efecto físico: mala iluminación, inclinación de la muestra, etc.

Los resultados obtenidos permiten descartar prácticamente esta herramienta dentro del análisis de imagen en muestras homogéneas con una insuficiente resolución espacial.

Como se indicó anteriormente, los métodos FCM son más viables, ya que se calcula el grado de pertenencia a diferentes grupos para cada píxel.

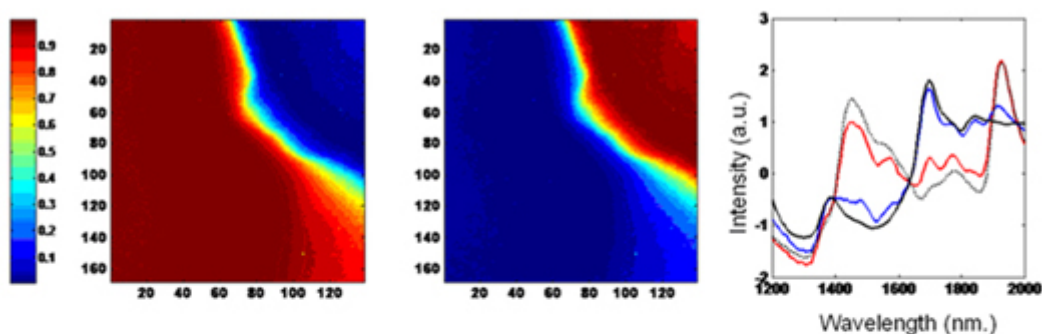
Considerando que cada muestra se compone de dos compuestos puros y un número indeterminado de combinaciones lineales entre ellos, y sabiendo que, al menos, hay dos clases claramente definidas, el análisis de FCM realizado teniendo en cuenta sólo dos grupos.

En la figura 5.25 se observan los clusters para 2 muestras. En las zonas que limitan analito puro y región de mezcla se observa una gradación en los colores, que permite tener una visión más real que la obtenida con la clasificación de KM Clustering. Los espectros puros y los centros calculados para cada cluster presentan coeficientes de correlación elevados entre sí.

RESULTADOS Y DISCUSIÓN

Para la muestra homogénea, se calculó el grado de pertenencia también con dos clusters obteniendo una distribución parecida a la obtenida con los anteriores métodos. De nuevo, los 2 centros muestran un coeficiente de correlación elevado entre si, por lo que no hay diferencias químicas entre los dos clusters, confirmando de nuevo que la variabilidad física afecta a los sistemas de registro de imagen.

MUESTRA 1



MUESTRA 4

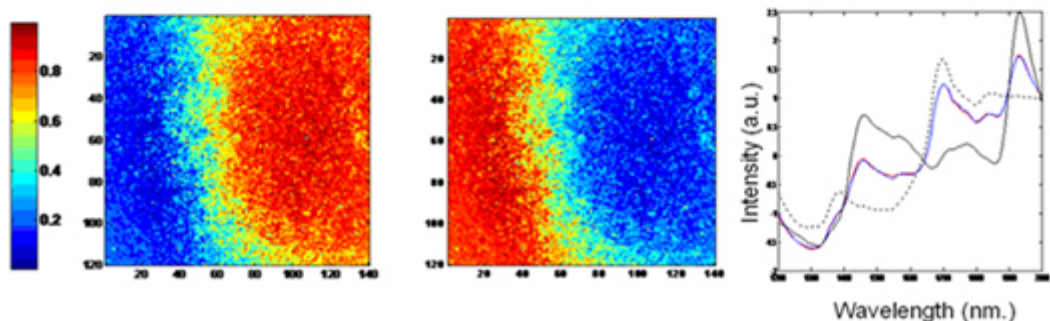


Figura 5.25. Representación de los FCM clustering obtenidos par alas muestras 1 y4.

A pesar que el FCM proporciona una información visual más cercana a la realizad, donde las regiones que limitan los analitos no pertenecen exclusivamente a un cluster, como sucedía con K Means, la capacidad discriminante en muestras homogéneas no ofrece resultados satisfactorios.

Análisis de distribución. Coeficientes de Correlación

RESULTADOS Y DISCUSIÓN

Como alternativa a los anteriores métodos de clasificación, e intentando mejorar los anteriores resultados, se presenta el coeficiente de correlación que permite, mediante un sencillo cálculo, proporcionar el coeficiente de correlación de cada píxel con los analitos. Con estos valores se pueden representar mapas de correlación para cada analito que permiten visualizar la aportación espectral de cada analito. La inspección visual de los mapas permite observar la distribución de los analitos en la superficie de la muestra. La representación de los histogramas de los coeficientes de correlación permite observar y comparar la distribución de los analitos en las muestras.

Los resultados obtenidos a partir de las imágenes de las tres muestras heterogéneas están representados en la Figura 5.26, donde los histogramas de los coeficientes de correlación permiten observar claramente la distribución de los analitos en la superficie de las muestras. De las tres muestras, los mapas de los coeficientes de correlación para el ibuprofeno, pueden distinguirse las zonas de mayor a menor correlación de acuerdo con la distribución espacial del analito en la muestra. En cuanto a los mapas de coeficiente de correlación en la mezcla de zonas, la intensidad de los cambios de color de la muestra 1 (color rojizo) a la muestra 3 (color verde), lo que indica que la correlación entre los espectros obtenidos en estas áreas con el puro del espectro ibuprofeno está disminuyendo de una muestra a otra en función de la proporción de ibuprofeno en la mezcla. En el otro lado, la correlación espectral de la región de mezcla con el almidón aumenta de la muestra 1 a la 3.

El coeficiente de correlación entre una imagen de ibuprofeno puro y una de almidón es de 0,3. En el histograma correspondiente al ibuprofeno de la muestra 1 (Figura 5.25), se pueden ver valores de coeficientes de correlación de píxeles que van desde el valor mínimo, que es 0,3, correspondientes a las regiones donde la presencia del ibuprofeno es nula o casi nula, hasta valores alrededor de 0,8 donde hay un número más alto de píxeles correspondientes a la zona de mezcla. Sin embargo la gran mayoría de píxeles presentan valores de coeficiente de correlación muy cercanos a 1. Para el almidón, el histograma muestra un comportamiento análogo, donde hay una mayoría de píxeles

RESULTADOS Y DISCUSIÓN

con valores de coeficiente de correlación de 0.3 correspondientes a la zona de ibuprofeno puro. Esto sugiere que existe una relación entre los coeficientes de correlación y la concentración de los analitos en las mezclas.

Los resultados para la muestra 1 son prácticamente análogos a los obtenidos en la muestra 3 debido a que la proporción de los analitos en la muestra 3 es idéntica cambiando de orden los analitos.

Comparando los histogramas para la muestra 2, con la misma cantidad de ibuprofeno y el almidón en la región de mezcla, se observa que el grupo de los coeficientes relacionados con la mezcla la zona de ibuprofeno aparece alrededor de 0,70, mientras que el grupo para el almidón aparece alrededor de 0,90. Esto indica que la correlación entre los espectros de la muestra y el espectro de almidón puro es superior a la correlación con el espectro de puro ibuprofeno. También se puede observar en las tonalidades de intensidad de colores de los mapas de correlación donde el mapa correspondiente al almidón presenta un color más intenso relacionado con coeficientes de correlación más elevados que para el mapa de Ibuprofeno. Este resultado puede atribuirse a la mayor absorción del almidón respecto al ibuprofeno y, en consecuencia, una mayor contribución a la señal espectral en la mezcla del almidón.

RESULTADOS Y DISCUSIÓN

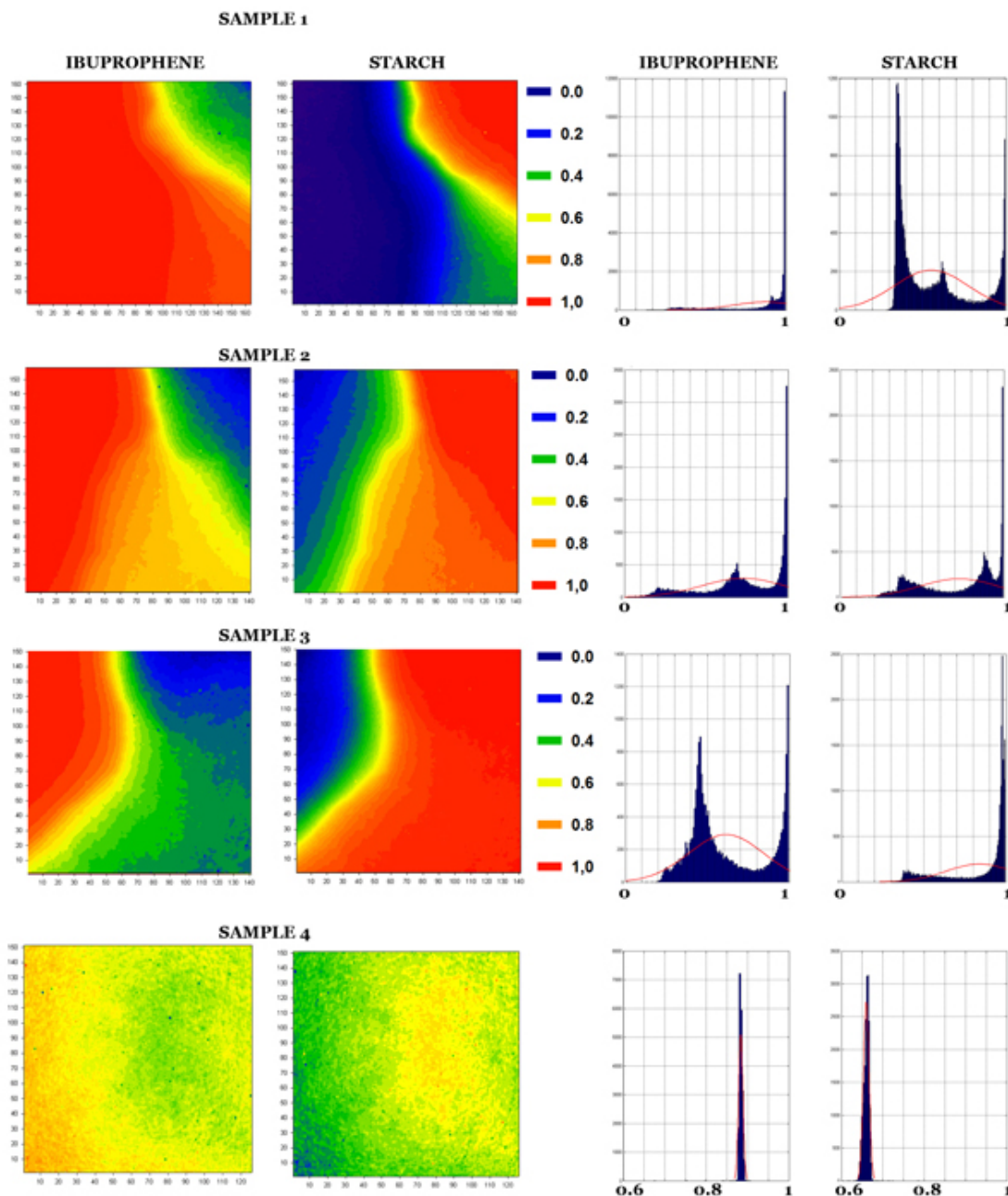


Figura 5.26. Mapas de coeficientes de correlación con los correspondientes histogramas. El color rojo en los mapas indica un alto grado de correlación mientras que el azul corresponde a regiones con poca correlación espectral.

Los resultados de la imagen de la mezcla homogénea (muestra 4) se muestran también en la figura 5.26. Los mapas de correlación obtenidos para los 2 analitos muestran una superficie plana que sugieren una homogeneización de los dos componentes en la muestra. Los histogramas de los coeficientes de correlación calculados entre la imagen de la muestra y los componentes puros encajan perfectamente en una campana de

RESULTADOS Y DISCUSIÓN

Gauss para ibuprofeno y el almidón, que ofrece un parámetro fiable cualitativo de la homogeneidad de la muestra. Comparando estos histogramas con los correspondientes de las muestras no homogéneas, la campana de Gauss podría ser una herramienta para confirmar la homogeneidad, además de la visualización de los mapas de la correlación. Estos resultados confirman el uso del coeficiente de correlación como una correcta alternativa a otros métodos de clasificación e identificación que no proporciona la información deseada.

5.3.2 ANÁLISIS CUANTITATIVO

Conocer la concentración de cada analito en cada píxel permitiría, además de calcular la concentración media de la muestra, visualizar mediante un mapa de concentraciones la distribución de los analitos en la muestra (información cualitativa)

En este apartado se presentan los resultados obtenidos con CLS y MCR-ALS. Una herramienta utilizada habitualmente es el PLS. Sin embargo, no se conoce la concentración de los analitos en cada píxel de la imagen, por lo que la calibración debería realizarse a partir de la concentración promedio de la imagen y no se aprovecharía la elevada cantidad de datos obtenidos con las técnicas de imagen.

CLS y MCR-ALS no requieren de una calibración previa ni conocer la concentración en cada píxel, a diferencia del PLS. Requieren, sin embargo, de los espectros puros de los analitos para realizar los cálculos y comparar la calidad de los mismos.

CLS

La figura 5.27 muestra los mapas de concentración obtenidos para los 2 analitos en una de las muestra heterogéneas, muestra 1, y en la muestra homogénea.

El interés del análisis cuantitativo está centrado en la región de mezcla, para las 3 primeras muestras, y en toda la superficies para la muestra 4. Para calcular la concentración se escogió un cuadro de 30 x 30 píxeles situado en la región de la mezcla de analitos.

RESULTADOS Y DISCUSIÓN

Los resultados se muestran en la tabla 5.10, donde se observan valores obtenidos alejados de los teóricos, sobretodo para el ibuprofeno en la muestra 2 y para el almidón en las muestras 1 y 2. Además la suma de concentraciones es superior en todas las muestras excepto la muestra 2 donde la suma es 90.

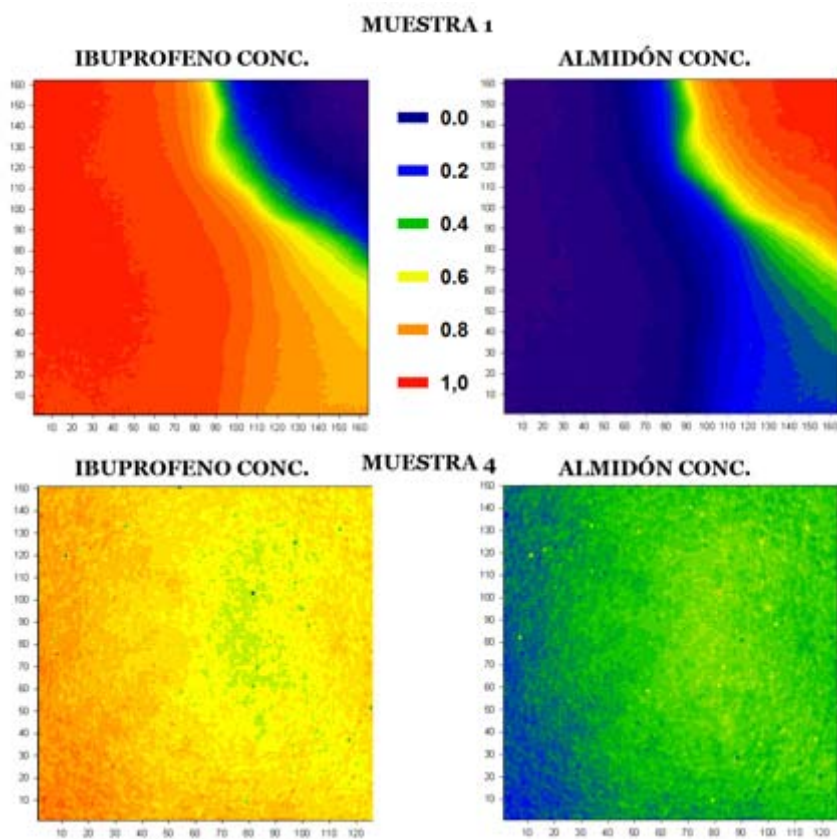


Figura 5.27. Mapas de concentración obtenidos por modelos CLS para las muestras 1 y 4.

Las concentraciones medias obtenidas para ibuprofeno y el almidón en la mezcla homogénea de 65% y 48%, respectivamente. El cálculo de la concentración de ibuprofeno coincide perfectamente con la teórica, pero la concentración de almidón está lejos de ser la esperada.

Tabla 5.10 Tabla de concentraciones obtenidas por CLS. Entre paréntesis se muestran los valores teóricos.

Muestra	Concentración CLS (Concentración real)	
	Ibuprofeno	Almidón

RESULTADOS Y DISCUSIÓN

1	79 (75)	40 (25)
2	48 (50)	74 (50)
3	15 (25)	70 (75)
4	65 (65)	48 (35)

Los resultados desfavorables obtenidos al aplicar esta herramienta puede ser atribuida al hecho de que los perfiles de los espectros obtenidos se ven afectados por otras fuentes de variabilidad externas a los dos analitos y por lo tanto presentan diferencias significativas con los espectros puros. Este es uno de los principales inconvenientes de la aplicabilidad de CLS en técnicas de imagen.

MCR-ALS Y VERSIÓN AUMENTADA

A diferencia del sistema directo de cálculo de la herramienta CLS, donde a partir del espectro puro y el espectro de cada píxel se han calculado las concentraciones en toda la imagen, el algoritmo MCR-ALS es un cálculo iterativo con infinitas soluciones y que permite incorporar información adicional que ayuden a resolver el cálculo iterativo.

La figura 5.28 muestra los resultados obtenidos de la muestra 1 (a) y de la muestra 4 (b). La descomposición MCR-ALS ha sido aplicada también a las muestras 2 y 3. Sin embargo, no se han representado los mapas de concentración de éstos 2. En la tabla 6 se muestran los valores obtenidos para la determinación de los 2 analitos en las 4 muestras. Además de los valores de cuantificación, se han representado los perfiles espectrales obtenidos para cada componente junto con los espectros puros de Ibuprofeno y almidón, y se han calculado, para cada muestra, el coeficiente de correlación entre los perfiles espectrales y los espectros puros, obteniendo información acerca de la calidad de la descomposición.

El estudio, donde se aplicó un Análisis por Componentes Principales sirve de punto de partida para el actual estudio. 2 componentes principales eran suficientes para aportar

RESULTADOS Y DISCUSIÓN

más de un 99 % de la variabilidad del sistema, por lo que puede presuponerse este número de factores como el inicial.

Utilizando los mismos pretratamientos que en los anteriores estudios se realizó la descomposición MCR aplicando como restricciones la no-negatividad de concentración y sistema cerrado (closure). Además como estimaciones iniciales se han aportado los espectros puros de Ibuprofeno y Almidón. Para las muestras heterogéneas los modelos con 2 componentes ofrecen resultados tanto cuantitativos como cualitativos correctos (tabla 5.11). Como se observa en la figura 5.27 (a), los mapas de concentración se corresponden con los mapas esperados teóricos y los perfiles espectrales obtenidos y los espectros puros, presentan coeficientes de correlación se cercanos a 1, confirmando la correcta descomposición. Dentro de los valores cuantitativos, los parámetros obtenidos para el almidón son, en las 3 muestras, superiores al valor teórico. Esto puede ser debido a la mayor contribución espectral por parte del almidón a la mezcla de los dos analitos como ya se vio en el apartado de coeficientes de correlación.

Tabla 5.11 Tabla de concentraciones obtenidas por MCR-ALS para las 4 muestras y la versión aumentada para la muestra 4. Entre paréntesis se muestran los valores teóricos.

Muestra	Concentración MCR-ALS		Concentración MCR-ALS Aumentado	
	(Concentración real)		(Concentración real)	
	Ibuprofeno	Almidón	Ibuprofeno	Almidón
1	73 (75)	26 (25)	--	--
2	45 (50)	54 (50)	--	--
3	23 (25)	77 (75)	--	--
4	57 (65)	43 (35)	65 (65)	35 (35)

RESULTADOS Y DISCUSIÓN

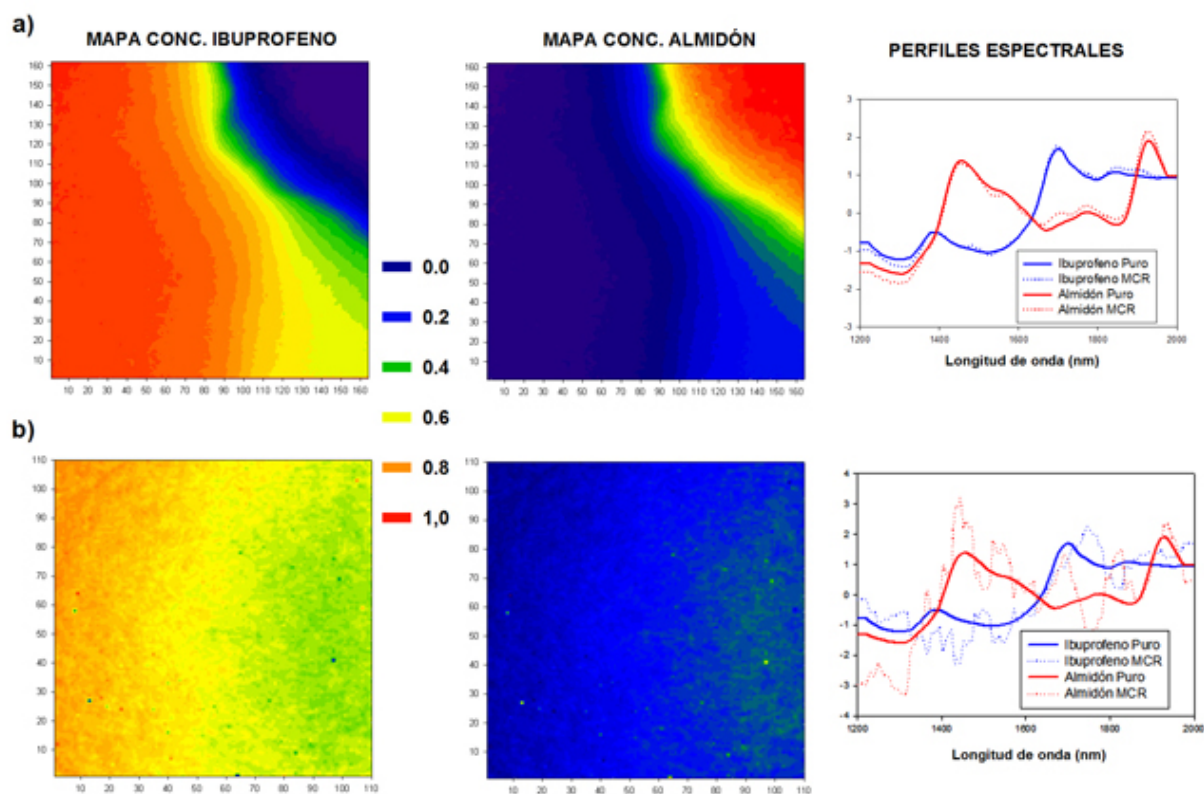


Figura 5.28. Mapas de concentración obtenidos por modelos MCR-ALS para las muestras 1 y 4. Se han representado los perfiles espectrales obtenidos en los componentes MCR y se han comparado con los espectros puros de Almidón (rojo) e Ibuprofeno (azul).

Para la muestra homogénea la descomposición de la imagen con 2 componentes no supera el 90% de la variabilidad. Los valores cuantitativos obtenidos para los 2 analitos en toda la muestra se alejan del valor teórico (57 % y 43% cuando la teórica es 65% y 35%). Los perfiles espectrales muestran un perfil muy ruidoso, alejado a los perfiles espectrales obtenidos para las muestras heterogéneas y su coeficiente de correlación con los espectros puros está alejado de 1. Los resultados cuantitativos y cualitativos confirman una incorrecta descomposición de la matriz. Se aumentó el número de componentes para la descomposición, para aumentar la variabilidad explicada por el modelo. De nuevo los resultados obtenidos no fueron buenos.

RESULTADOS Y DISCUSIÓN

Con el objetivo de mejorar la descomposición para la muestra homogénea se planteó un modelo MCR-ALS aumentado, que consiste en aumentar la matriz de información inicial, en ese caso con espectros puros de los dos analitos presentes en la mezcla.

Los nuevos resultados con 2 componentes y más de un 98 % de varianza se muestran en la figura 5.29, donde el valor promedio de concentración para los dos analitos coincide perfectamente con el valor esperado(tabla 5.11), mientras que los nuevos perfiles espectrales presentan un coeficiente de correlación con los espectros puros cercano a 1 .

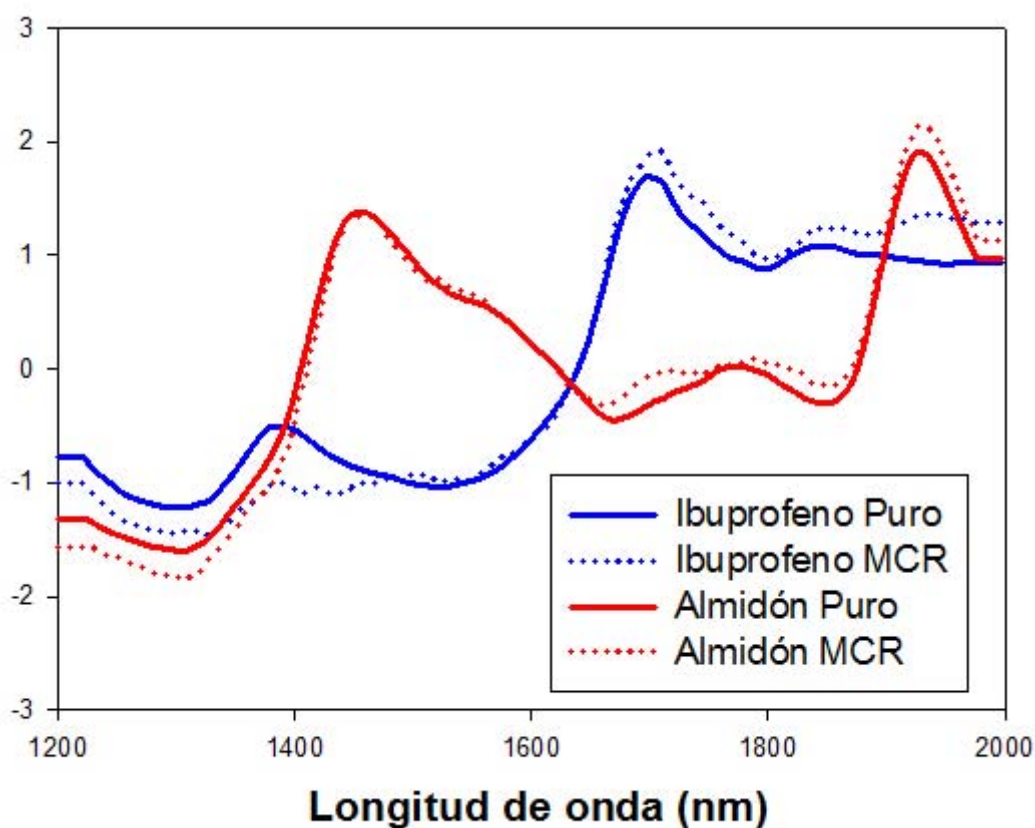


Figura 5.29. Perfiles espectrales obtenidos al aplicar el MCR-ALS aumentado.

Los resultados obtenidos con MCR-ALS para las muestras heterogéneas son correctos, tanto a nivel de concentraciones como a nivel cualitativo, donde los espectros obtenidos en el cálculo están altamente correlacionados con los espectros puros.

RESULTADOS Y DISCUSIÓN

A pesar que los resultados iniciales para la muestra homogénea no son correctos, el uso de la matriz aumentada ha solventado esas limitaciones, haciendo de esta, una técnica idónea incluso para el tratamiento de muestras homogéneas, y por lo tanto podría ser aplicada a muestras farmacéuticas reales.

6. ARTÍCULOS

ACEPTADOS PARA LA

TESIS DOCTORAL

**6.1 PREPARING CALIBRATION SETS FOR USE IN
PHARMACEUTICAL ANALYSIS BY NIR
SPECTROSCOPY.**

M. Blanco, M. Bautista, M. Alcalà
Journal of Pharmaceutical Sciences. Volume97, Issue 3,
2008. Pages: 1236-1245

Preparing Calibration Sets for Use in Pharmaceutical Analysis by NIR Spectroscopy

M. BLANCO, M. BAUTISTA, M. ALCALÁ

Department of Chemistry, Analytical Chemistry Unit, Faculty of Sciences, Autonomous University of Barcelona, E-08193 Bellaterra, Spain

Received 26 March 2007; revised 8 May 2007; accepted 28 May 2007

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/jps.21105

ABSTRACT: A new methodology for constructing calibration sets based on the use of laboratory samples encompassing the same variability sources as production samples was developed. The proposed methodology requires the use of no reference method in order to obtain reference values for the analyte; also, it provides more simple and robust calibration models than does the conventional methodology while retaining its predictive capacity. The procedure involves subjecting a set of laboratory samples spanning the desired API concentration range to a granulation treatment similar to that of the industrial process in order to obtain samples with the same physical variability as the production samples. The laboratory samples thus obtained are used to develop partial least squares (PLS1) calibration models in order to quantify the API in a pharmaceutical granulate. Based on the results obtained in this work, NIR spectroscopy is an effective alternative to the reference methods currently used for calibration. The proposed methodology requires no reference values to construct models; therefore, it can be regarded as an absolute analytical method. Also, it confirms the advantages of NIR spectroscopy as part of the process analytical technology (PAT) used by the pharmaceutical industry. A second aim has been the use of the multiplicative curve resolution-alternating least squares (MCR-ALS) algorithm to examine potential polymorphic transformations of the API during granulation. © 2007 Wiley-Liss, Inc. and the American Pharmacists Association *J Pharm Sci*

Keywords: NIR spectroscopy; API determination; PLS calibration; MCR-ALS

INTRODUCTION

Industrial pharmaceutical is a strictly activity regulated by various national and international organisms and subject to quality control procedures involving high numbers of physical and chemical determinations of both raw materials, and intermediate and end products. Pharmaceutical quality control procedures require the use of a wide range of analytical methodologies often involving labor-intensive, time-consuming analytical protocols that preclude the expeditious

obtainment of results. In addition, the development and implementation of process analytical technologies (PATs) for the overall control of production processes entails the development and application of simple, fast analytical techniques capable of delivering analytical information in an expeditious manner.

Near infrared spectroscopy (NIRS) has provided an effective solution to many problems in this context, as reflected in the large number of applications currently in use by the pharmaceutical industry^{1–4} and its ability for adaptation to virtually all types of samples with minimal intervention of the analyst.

The active principle ingredient (API) of a pharmaceutical formulation is usually determined by high performance liquid chromatography (HPLC),

Correspondence to: M. Blanco (Telephone: +34 93 581 13 67; Fax: +34 93 581 13 67; E-mail: marcel.blanco@uab.es)

Journal of Pharmaceutical Sciences
© 2007 Wiley-Liss, Inc. and the American Pharmacists Association

which is time-consuming and destructive, and produces waste. Determining physical properties of formulations entails using other analytical techniques that are often slow, labor-intensive and destructive.

The NIRS technique affords the simultaneous determination of physical and chemical parameters in a rapid, clean, non-destructive manner;⁵⁻⁷ also, it allows some physical properties of the API to be monitored throughout a pharmaceutical production process.

The determination of chemical and physical properties by NIR spectroscopy involves the use of multivariate calibration models. The samples to be used to construct such models should encompass every source of variability in the process potentially influencing the properties of the end product. Samples of the end product are unsuitable for this purpose as they span too narrow a chemical composition range; this entails expanding the calibration model with laboratory samples of appropriate composition in order to develop accurate, precise and robust calibration models.⁸ The use of production samples to construct PLS calibration models entails the use of a reference method in order to determine the parameter to be employed to construct the model.

Some APIs used in pharmaceuticals exist as several polymorphs that must be identified prior to inclusion in a production process. However, drug manufacturing processes can cause polymorphic changes in APIs and alter the final properties of the formulation.^{9,10} Recently, some authors have in fact reported polymorphic changes caused by granulation or compaction during drug production processes.¹¹⁻¹³

In this work, we developed a new methodology based on the use of laboratory samples encompassing the same variability sources as the production samples in order to reduce the complexity of conventionally constructed models while preserving their robustness, all without the need of a reference method. The proposed methodology involves subjecting a set of laboratory samples spanning the desired API content range to a granulation procedure similar to the industrial procedure in order to obtain samples encompassing the same physical variability as the production samples. In addition, we examined the potential polymorphic transformation of the API during granulation; in fact, dexketoprofen has been reported to undergo such a transformation during its wet granulation in the presence of microcrystalline cellulose as excipient.¹⁴

EXPERIMENTAL METHODOLOGY

Reagents

The pharmaceutical formulation studied was a solid granulate containing 10 mg/g dexketoprofen as active principle ingredient (API), 970 mg/g sucrose (Suc) as major excipient and other, minor excipients including lemon flavor (Fla), neohesperidin dihydrochalcone (Neo), Quinoline Yellow (Qui) and ammonium glycyrrhizinate (Gly). The production samples and excipients used were kindly supplied by Laboratorios Menarini, S.A. (Badalona, Spain).

Samples

Laboratory Samples for Determining the API

A total of 16 placebos consisting of the five excipients in the formulation at concentrations spanning a range $\pm 10\%$ around each nominal value were prepared.

The laboratory samples were obtained by mixing appropriate amounts of active principle and placebo. To this end, aliquots of the placebo were randomly selected and mixed with variable amounts of API in order to prepare 45 laboratory samples spanning an API content range $\pm 30\%$ around the nominal value.

Also, 5 g aliquots of each of the 45 previous samples were granulated following spraying with 1.5–3.0 g of water. The wet paste thus obtained was mixed by hand in a glass mortar with the aid of a spatule for 20 min, which produced a mass of granules of appropriate size that was dried in a Vaciotem stove from P-Selecta (Abrera, Spain) at 60°C for 5 h. Once dry, the granulate was ground by hand to a particle size similar to that of the commercial product. The resulting product was used as such to record its NIR spectrum.

The API content in each sample were calculated from the values obtained by weighing.

Production Samples

A total of 30 samples were withdrawn from their production line over a period of 6 months. Each sample was from a different batch and its API content determined using the HPLC method for this purpose previously developed and validated by Laboratorios Menarini.

Samples for Examining Polymorphic Transformations

Polymorphic changes in the API were studied by preparing seven binary samples containing sucrose and the API at concentrations from 0 to 20 mg/g. Aliquots of each sample were subjected to the above-described granulation procedure and their NIR spectra recorded prior to and after granulation.

Recording of NIR Spectra

Near infrared spectra were recorded on a Model 5000 NIR spectrophotometer from FOSS NIRSystems (Silver Springs, MD) equipped with a Rapid Content Analyzer (RCA) module and governed via the software Vision 2.51.

A reference spectrum was recorded prior to each analytical spectrum by using a ceramic plate. The NIR spectra for the laboratory samples were recorded both prior to (powder) and after granulation (granules). To this end, aliquots were placed in a quartz cell and their reflectance spectra recorded in triplicate over the wavelength range 1100–2498 nm, using a spectral resolution of 2 nm and 32 scans per spectrum. Samples were turned over between successive recordings. The average spectrum obtained from the three replicates for each sample was subsequently used to construct and validate the calibration models, and also to quantify the API in the pharmaceutical preparation. The spectra for the production samples were recorded in the same manner, without the need for any pretreatment.

Processing of NIR Data

The API content of the production samples was determined by using PLS1 models constructed from various calibration sets, using the chemometric software The Unscrambler v. 9.2, from Camo Process AS (Oslo, Norway).

The spectral treatments used to construct the calibration models included the Standard Normal Variate (SNV), and the first and second derivatives. Derivative spectra were obtained by using the Savitzky–Golay algorithm with a 11-point moving window and a second-order polynomial. Calibration models were constructed by cross-validation. The optimum number of PLS-Factors was chosen to be that corresponding to the minimum of a plot of PRESS as a function of

the number of PLS components, the former being defined as

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

The calibration and prediction quality was assessed in terms of the Relative Standard Error (RSEC for calibration and RSEP for prediction):

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

The spectrum for the API was obtained by using the multivariate curve resolution-alternating least squares (MCR-ALS) algorithm, which was implemented by running the software GUIPRO (Gemperline and Cash, East Carolina University) under the MATLAB software following wet granulation of the product. MCR-ALS deconvolution provides the spectral and concentration profiles for the ingredients of an evolving process. This algorithm has no unique solution, so the results must be subjected to natural constraints such as closure, non-negativity and unimodality, and also to equality constraints (analyte spectra and concentrations).^{15,16}

The treatments used to obtain the pure spectrum for the API included the standard normal variate (SNV), and the Savitzky–Golay first and second derivative.

The goodness of the results was assessed by comparing the MCR spectra for each species with the corresponding reference standard via a correlation coefficient.

RESULTS AND DISCUSSION

Constructing an accurate, precise, robust calibration model entails using a calibration model including every source of physical and chemical variability in the set of samples to be determined. Chemical (composition) variability is usually introduced by including in the calibration set samples spanning a wide range around the nominal value in order to ensure the accurate determination of any production sample. Particle size, its distribution, and the specific analyte polymorph are among the physical variability sources to be included in the calibration model.

Table 1. Spectral Correlation between the Components of the Pharmaceutical Formulation

	API	Qui	Neo	Gly	Fla	Suc
API	1					
Qui	0.878	1				
Neo	0.936	0.950	1			
Gly	0.969	0.916	0.979	1		
Fla	0.937	0.919	0.990	0.979	1	
Suc	0.883	0.778	0.920	0.922	0.941	1

Commercial samples invariably span too narrow an API content range (usually $\pm 5\%$ around the nominal value) and are thus unsuitable for constructing calibration models. This entails using laboratory samples in order to expand such a range. The high spectral correlation between components in our formulation (see Tab. 1) required reducing correlation between their concentrations in order to obtain robust models. To this end, placebos spanning a concentration range $\pm 10\%$ around the nominal value and exhibiting a low correlation between components were prepared; however, two excipients only (*viz.* Suc and Fla) accounted for 99.8% of the total composition of the placebo between them, which inevitably resulted in a higher correlation between their concentrations. An error in dosing the excipients during the

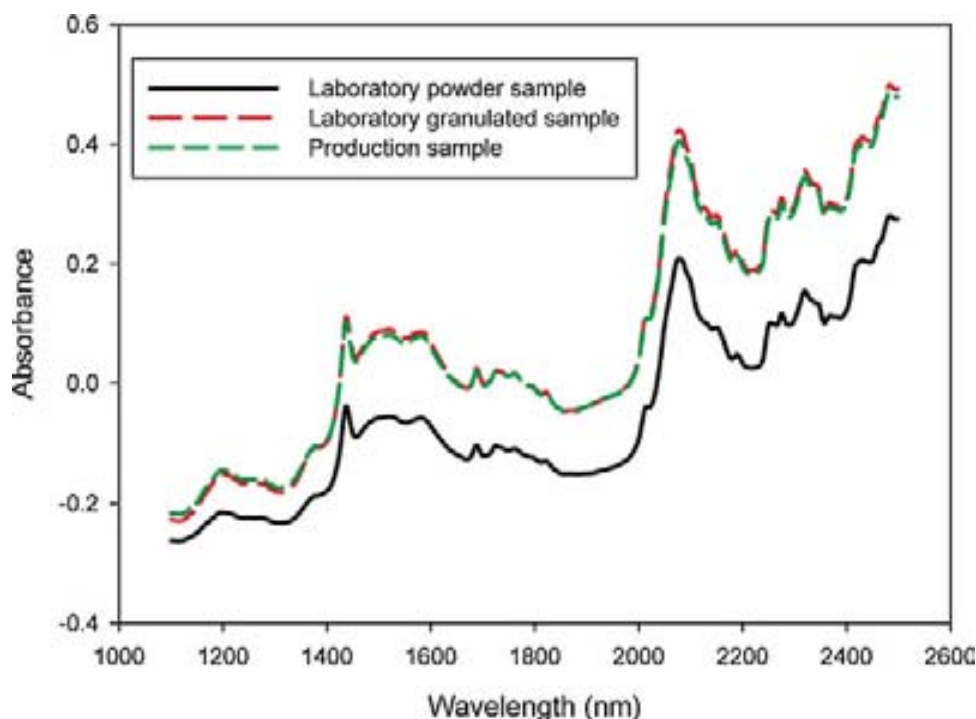
Table 2. Correlation between the Concentrations of the Different Components in 45 Samples

	API	Qui	Neo	Gly	Fla	Suc
API	1					
Qui	0.065	1				
Neo	0.092	0.522	1			
Gly	0.284	0.172	0.304	1		
Fla	0.173	0.425	0.387	0.120	1	
Suc	-0.329	-0.436	-0.411	-0.181	-0.986	1

industrial preparation process can therefore lead to gross errors in predicting the API content by effect of its correlation with some component of the placebo. This should be avoided by preparing appropriate laboratory samples. We used a total of 16 placebos that were selected at random for mixing with variable amounts of API in order to prepare 45 samples with a view to constructing the calibration and validation sets.

Table 2 shows the coefficients of correlation between the different components in the 45 samples. As can be seen, the highest coefficient of the API with any excipient was 0.329.

These laboratory powder samples spanned the whole range of chemical variability potentially encompassed by the commercial product and were granulated as described in the Experimental

**Figure 1.** NIR spectra for laboratory powder, laboratory granulated and production samples. The spectral differences are due to differences in particle size between sample.

section to obtain a second sample set additionally incorporating the physical variability introduced by granulation—the particle size obtained was similar to that of the commercial product.

Figure 1 shows the absorbance spectrum for a production sample, a laboratory powder sample and a laboratory granulated sample with API content close to the nominal value. As can be seen, the two spectra are shifted with respect to the other one by effect of the differences in particle. Such an effect is also apparent in the score scatter plot obtained from a PCA of the absorbance spectra for laboratory powder samples and production samples (Fig. 2). The second score (1.5% X-variance) essentially reflects differences in API concentrations. The score scatter plots obtained upon spectral treatment (*viz.* SNV or derivation) exhibit a similar scatter; this indicates that the treatments fail to completely suppress scattering, even though they reduce it—which is not apparent in the graphs.

We tested various PLS1 models in order to quantify the API in production samples; to this end, we used various calibration sets, spectral ranges and treatments. Roughly two-thirds of the 45 laboratory samples were included in the calibration set and one-third in the validation set. The

samples in the calibration set were assumed to encompass the whole variability of the production process and the required concentration range. All others were used to validate the model.

Table 1 shows the figures of merit of the model and the predictions obtained for laboratory powder samples. As can be seen, the second-derivative spectral treatment (model 1) was that providing the best results. The calibration set consisted of 27 samples and the validation set of 18. Using 5 PLS factors ensured correct calibration (98.5% Y-explained variance) and acceptable predictive capacity (RSEC = 2.4%) for laboratory powder samples, but not for laboratory granulated and production samples; in fact, the results for the latter two were subject to gross negative systematic errors that were ascribed to differences in physical properties (particle size) between the two types of samples.

Including production samples in the calibration set allowed physical variability to be introduced and the predictions for laboratory granulated and production samples to be improved. Model 2 was constructed by adding seven production samples to the calibration set for model 1. The new model required two more PLS factors than the previous one by effect of the inclusion of samples with

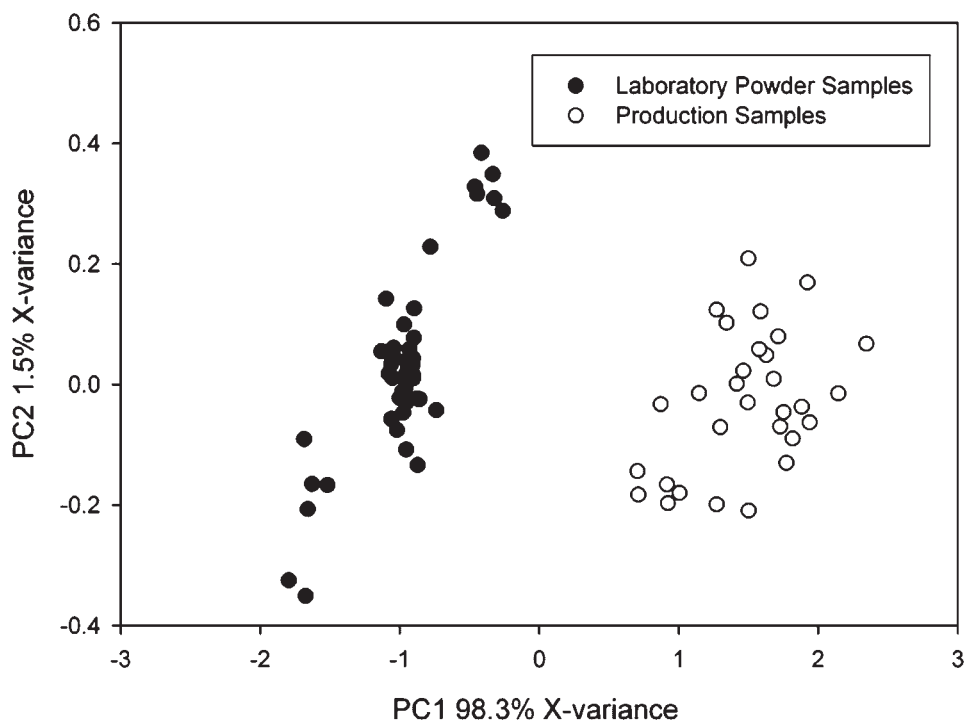


Figure 2. PCA scatter plot constructed from absorbance spectra over the wavelength range 1100–2498 nm for laboratory powder and production samples. The differences in particle size are related to the distribution of scores in the first PC.

spectral differences; also, the predictions for production samples were better, but still inadequate—RSEP values were below the required levels.

In order to improve the results provided by the previous two models, we conducted similar tests but using laboratory granulated samples. Figure 1 shows the absorbance spectra for the same laboratory sample following granulation. The slight differences in particle size between the two types of samples are consistent with their spectral similarity. Based on the scatter plot for the first two scores for both types of samples (Fig. 3), spectral differences between them are virtually zero. All samples clustered together, which suggests that they were subject to the same sources of spectral variability. PC1 (95.5%) was associated to the API concentration. The production samples gathered in the center of the cluster by effect of the narrow API concentration range they spanned ($\pm 5\%$). On the other hand, the laboratory granulated samples spanned a wider API content range and resulted in a broader distribution of scores along the PC1 axis.

We constructed a third calibration model (model 3) based on laboratory granulated samples alone (25 for calibration and 20 for validation). As can be

seen, the calibration statistics were quite acceptable: 5 PLS factors, RSEC = 2%, and an RSEP value of 3.9% and 1.6% for laboratory granulated and production samples, respectively. This model features the simplicity of model 1 as regards the number of PLS models and improves on the predictive capacity of model 2 (Tab. 3).

Figure 4 shows the residuals of the predictions of the three models for the production samples. As can be seen, model 1 exhibited very high systematic residuals by effect of its excluding physical variability in the production process. On the other hand, the residuals of model 2 oscillated randomly around 0, but were close to 20% in some instances. Finally, the residuals of model 3 were very close to 0 and randomly distributed.

Polymorphic Transformation

Dexketoprofen was recently showed¹⁴ to undergo a polymorphic transformation into an amorphous form during wet granulation in the presence of microcrystalline cellulose. The similarity of the granulation process led us to investigate whether such a transformation might also take place during our tests.

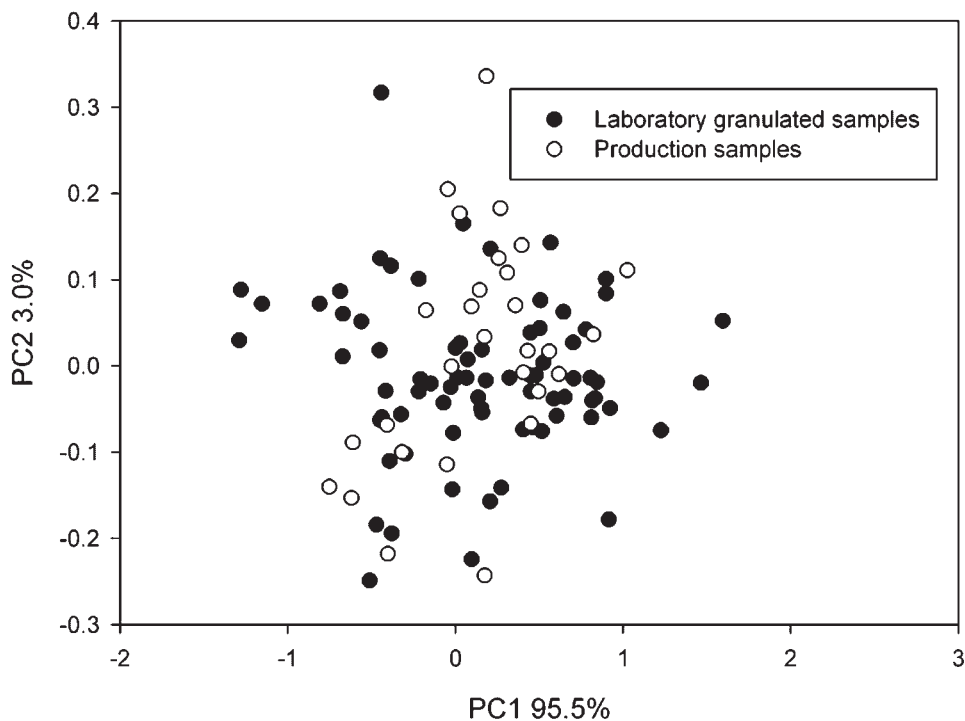


Figure 3. PCA scatter plot constructed from absorbance spectra over the wavelength range 1100–2498 nm for laboratory granulated and production samples. The differences in concentration are related to the first PC.

Dexketoprofen occurs in two polymorphic crystalline forms (A and B). In this work, we used polymorph B. We applied the MCR-ALS algorithm to powder and granulated binary samples containing the API and sucrose in order to obtain the spectrum for the API, using the above-described procedure. We used the entire wavelength range in conjunction with the absorbance, SNV and first- and second-derivative spectral modes, and applied concentration non-negativity and closure constraints; the latter are applicable to closed systems and assume the combined concentrations of all chemical species to be constant throughout. The previous information was supplemented with the sucrose concentration in each sample and its spectrum. An identical process was applied to both laboratory powder and granulated samples.

The API spectra obtained for the powder and granulated samples were compared with those for polymorph B of the API and its amorphous counterpart. Figure 5 shows the 2250–2350 nm region in the second-derivative spectrum that clearly exposes the differences between the two polymorphs. As can be seen, the MCR-ALS spectra for the powder and granulated samples were very similar to that for the pure API. Such a similarity was confirmed by the correlation coefficients for the spectra (Tab. 4) over the wavelength range 2250–2350 nm. Also, the whole API spectra obtained for the powder and granulated samples were compared between them in Absorbance mode and correlation between these two spectra was 0.99.

By contrast, correlation with the spectrum for the amorphous form was much lower, which suggests that it undergoes no polymorphic transformation during granulation.

CONCLUSIONS

The proposed methodology for preparing calibration samples incorporating the intrinsic variability of the production process allows more simple models of acceptable predictive capacity to be constructed without the need for any production samples.

Reference values for the API content in the samples can be established from the weights of the components in laboratory-made mixtures; this dispenses with the need to use a reference method for their determination, which is the usual practice in constructing NIR calibration models. Therefore, the proposed NIR methodology provides an absolute method of analysis as it requires no

Table 3. Figures of Merit of the PLS Models Constructed from Laboratory Powder and Granulated Samples, and Production Samples

Model	Calibration Samples	Spectral Pretreatment	Spectral Range	PLS Factors	Calibration			Prediction							
					Y Explained Variance %	RSEC%	RSEP%	Laboratory			Production				
								Powder	Granulated	Residuals	Powder	Granulated	Residuals		
1	Laboratory powder	Second derivative	1100–1200; 1600–2500	5	98.5	2.4	5.6	18	66.5	45	44.8	30	–3.97	1.81	12.01 ^a
2	Laboratory powder and production	Second derivative	1100–2500	7	96.7	3.2	8.6	18	46.2	45	8.5	23	0.23	0.82	1.51 ^b
3	Laboratory granulated	Second derivative	1100–1200; 1400–2500	5	98.7	2.0	—	—	3.9	20	1.6	30	–0.05	0.13	2.00 ^a

^at crit = 2.04 (95%, n = 30).

^bt crit = 2.07 (95%, n = 23).

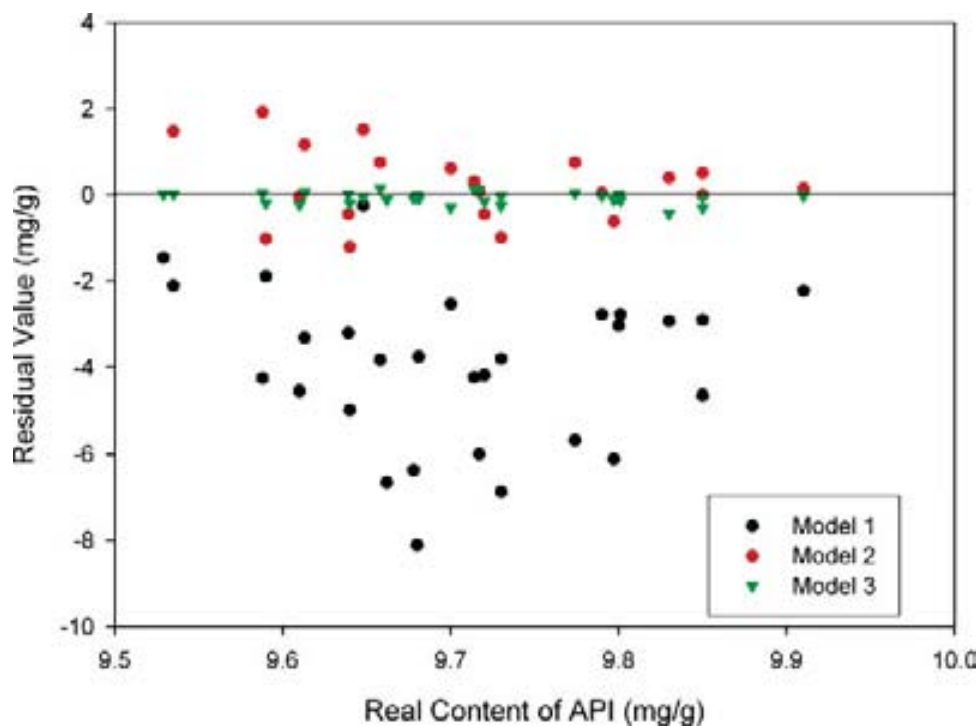


Figure 4. Residual values obtained in the prediction of production samples.

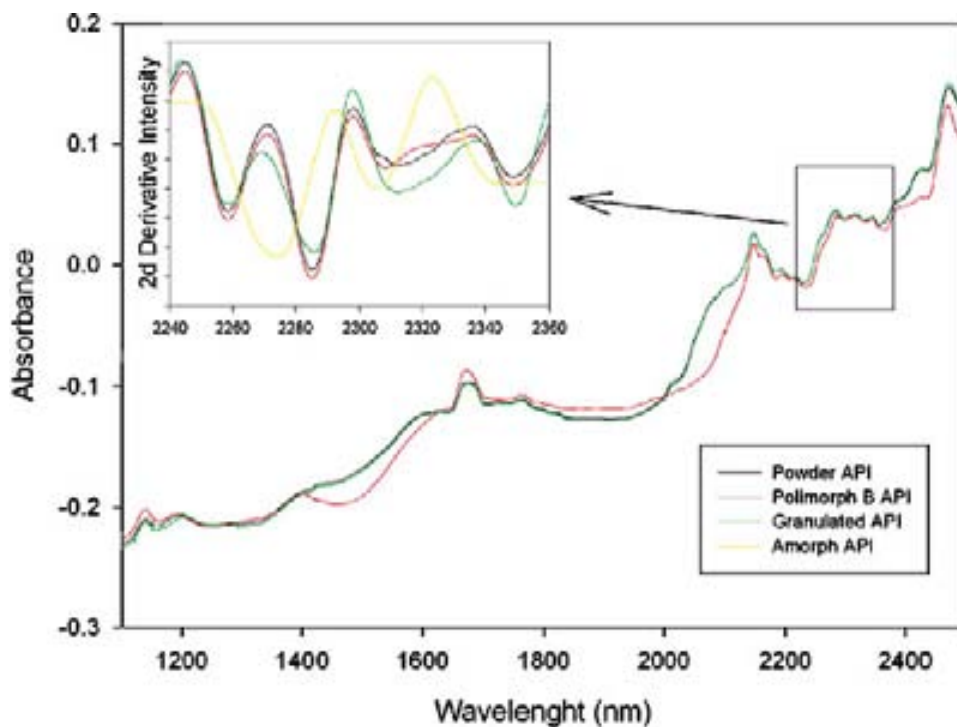


Figure 5. NIR spectra obtained over the wavelength range 2250–2350 nm for polymorph B of the API, powder and granulated APIs (with MCR-ALS) and the amorphous API where differences between the amorphous API and the other three forms are clearly apparent.

Table 4. Coefficients of Correlation between the Second-Derivative Spectra for Different API Forms in the Spectral Region 2250–2350 nm

	Pure API	Granulated API	Powder API	Amorphous API
Polymorph B API	1			
Granulated API	0.909	1		
Powder API	0.999	0.909	1	
Amorphous API	0.294	0.283	0.294	1

reference method to analyze the samples used to construct the calibration model.

Application of the MCR-ALS algorithm to mixtures of dexketoprofen and sucrose revealed that the former undergoes no polymorphic transformation during granulation, but it does in the presence of cellulose. Therefore, whether or not such a transformation occurs depends on the nature of the support used.

ACKNOWLEDGMENTS

The authors are grateful to Spain's MCyT for funding this research within the framework of Project BQU2003-04247. Also, they wish to thank Laboratorios Menarini, S.A. (Badalona, Spain) for kindly supplying the studied samples.

REFERENCES

- Blanco M, Alcalá M, González JM, 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* 95:2137–2144.
- Blanco M, Alcalá M. 2006. Simultaneous quantitation of five active principles in a pharmaceutical preparation: Development and validation of a near infrared spectroscopic method. *Eur J Pharm Sci* 27:280–286.
- Colón J, Peroza C, Caraballo W, Conde C, Li T, Morris KR, Románach RJ. 2005. On line non-destructive determination of drug content in moving tablets using near infrared spectroscopy. *Proc Anal Tech* 2:8–15.
- Paris I, Janoly-Dumenil A, Paci A, Mercier L, Bourget P, Brion F, Chaminade P, Rieutord A. 2006. Near infrared spectroscopy and process analytical technology to master the process of busulfan paediatric capsules in a university hospital. *J Pharmaceut Biomed Anal* 41:1171–1178.
- Skibsted ETS, Boelens HFM, Westerhuis JA, Witte DT, Smilde AK. 2006. Simple assessment of homogeneity in pharmaceutical mixing processes using a near-infrared reflectance probe and control charts. *J Pharmaceut Biomed Anal* 41:26–35.
- Blanco M, Alcalá M. 2006. Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy. A contribution to process analytical technologies. *Anal Chim Acta* 557:353–359.
- Feng YC, Chang-Qin H. 2006. Construction of universal quantitative models for determination of roxithromycin and erythromycin ethylsuccinate in tablets from different manufacturers using near infrared reflectance spectroscopy. *J Pharm Biomed Anal* 41:373–384.
- Blanco M, Alcalá M. 2005. Use of near-infrared spectroscopy for off-Line measurements in the pharmaceutical industry. In: Bakeev AK, editor. *Process analytical technology*. 1st edition. Oxford: Blackwell Publishing. Chapter 10, pp 375–376.
- Hancock B, Carlson G, Ladipo D, Langdon B, Mullarney M. 2002. Comparison of the mechanical properties of the crystalline and amorphous forms of a drug substance. *Int J Pharm* 241:73–85.
- Williams A, Cooper V, Thomas L, Griffith L, Petts C, Booth S. 2004. Evaluation of drug physical form during granulation, tableting and storage. *Int J Pharm* 275:29–39.
- Rasanen E, Rantanen J, Jorgensen A, Karjalainen M, Paakkari T, Yliruusi J. 2005. Novel identification of pseudopolymorphic changes of theophylline during wet granulation using near infrared spectroscopy. *J Pharm Sci* 90:389–396.
- Rantanen J, Wikström H, Turner R, Taylor L. 2001. Use of in-line near-infrared spectroscopy in combination with chemometrics for improved understanding of pharmaceutical processes. *Anal Chem* 77:556–563.
- Li W, Worosila G, Wang W, Mascaro T. 2005. Determination of polymorph conversion of an active pharmaceutical ingredient in wet granulation using NIR calibration models generated from the premix blends. *J Pharm Sci* 94:2800–2806.

14. Blanco M, Alcalá M, González JM, Torras E. 2006. Near infrared spectroscopy in the study of polymorphic transformations. *Anal Chim Acta* 567: 262–268.
15. Tauler R, Kowalsky B. 1993. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal Chem* 65:2040–2047.
16. de Juan A, Tauler R. 2003. Chemometrics applied to unravel multicomponent processes and mixtures. Revisiting latest trends in multivariate resolution. *Anal Chim Acta* 500:195–210.

**6.2 PHARMACEUTICAL GEL ANALYSIS BY NIR
SPECTROSCOPY. DETERMINATION OF THE
ACTIVE PRINCIPLE AND LOW CONCENTRATION
OF PRESERVATIVES.**

M. Blanco, M. Alcalà, M. Bautista
European Journal of Pharmaceutical Sciences. 33 (2008)
Pages: 409–414

available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/ejps

Pharmaceutical gel analysis by NIR spectroscopy Determination of the active principle and low concentration of preservatives

M. Blanco*, M. Alcalá, M. Bautista

Departament de Química, Unitat de Química Analítica, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain

ARTICLE INFO

Article history:

Received 4 September 2007

Received in revised form

8 January 2008

Accepted 22 January 2008

Published on line 2 February 2008

Keywords:

Near infrared spectroscopy

Pharmaceutical preparation

Active principle

Preservatives

PLS

Validation

ABSTRACT

Near infrared spectroscopy has proved highly suitable for the analysis of pharmaceutical formulations. However, its limited sensitivity can severely restrict its scope of application. In this work, we determine the active principle and two preservatives in a pharmaceutical preparation available as a hydrogel. The matrix of the pharmaceutical preparation exhibits strong absorption in the NIR spectral region; also, the two preservatives (parabens) are very similar in chemical and spectral terms, and present at low concentrations in the pharmaceutical. These complications make it rather difficult to accurately quantify the active principle and the preservatives, which can only be accomplished by using an effective design in order to prepare the samples to be included in the calibration set and select the optimum spectral range for measuring each analyte. The evaporation of solvents during the measurement process produces increasing errors related with sample's air exposition; the introductions of new samples with a wider range of the volatile components correct this effect. An ANOVA of the predictions obtained with the new models shows that correct the error due to evaporation. The proposed method was validated for the analytical control of the studied preparation.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

A sizeable fraction of pharmaceutical formulations is available in the form of liquid or semi-liquid (gel) preparations consisting of solutions or suspensions of one or several active principles in water or a hydroalcoholic mixture as solvent. In addition, many preparations contain some preservative intended to ensure that the formulation will retain its pharmacological properties for as long as required. Parabens, which are methyl, ethyl or propyl esters of *p*-hydroxybenzoic acid, are among the most widely used pharmaceutical preservatives. The members of this compound family have the advantage that their behaviour is virtually independent of pH over a wide

range. Parabens are effective preservatives against moulds and yeasts, but much less so against bacteria. Their effectiveness – and also their insolubility – increases with increasing chain length; their specific effects, however, depend on the particular microorganism (<http://milksci.unizar.es/bioquimica/temas/aditivos/conservantes.html>). Usually, parabens are used in mixtures and at low concentrations in order to avoid any toxic effects. Regulations currently in force require their determination in order to ensure effective preservation of the formulations containing them.

The methods usually employed for the analytical control of pharmaceutical preparations are of the HPLC type; also, their application involve labour-intensive, time-consuming proce-

* Corresponding author. Tel.: +34 93 581 1367; fax: +34 93 581 1367.

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

0928-0987/\$ – see front matter © 2008 Elsevier B.V. All rights reserved.

doi:10.1016/j.ejps.2008.01.011

dures and the use of reagents and solvents (Jing et al., 2006; Quaglia et al., 2005) and consequently are not appropriate to quality control. The delay in obtaining the analytical results precludes taking timely decisions as regards corrective actions on the process. Near infrared spectroscopy has previously been successfully used to develop fast, accurate, precise analytical methods for the pharmaceutical industry (Colón et al., 2005; Blanco and Alcalá, 2006). Though the preparation of a NIR method is lightly more labourious than HPLC, its application presents many advantages over HPLC as the low time of analysis, no pre-treatment of sample and not need of specialized staff. The flexibility of this technique and its associated computational procedures allow virtually all types of industrial analyses to be undertaken. Also, NIRS methods require virtually no sample treatment, which facilitates the obtainment of analytical results within a few seconds and with minimal intervention of the analyst.

However, the limited sensitivity of the NIRS technique can hinder its use for the simultaneous determination of the active principle and two preservatives in a pharmaceutical preparation in gel form where the active principle is present at a fairly high concentration whereas the two preservatives are in very small amounts. In this work, we explored the combined use of NIRS and partial least-squares (PLS) multivariate calibration models with a view to quantifying the analytes at the low concentrations present in the pharmaceutical.

2. Materials and methods

2.1. Samples

The pharmaceutical formulation studied consisted of a hydroalcoholic gel containing ketoprofen (KET, 25 mg g⁻¹) as active principle, methyl *p*-hydroxybenzoate (MET, 0.8 mg g⁻¹) and propyl *p*-hydroxybenzoate (PRO, 0.2 mg g⁻¹) as preservatives, and water and ethanol (more than 950 mg g⁻¹ between the two) as major excipients. A total of 30 samples from 10 different production batches were analysed.

A D-optimal experimental design was used to prepare 33 laboratory samples in order to expand the ketoprofen and preservative concentration ranges by $\pm 50\%$ around their respective nominal contents in the formulation. The three ingredients were added in such a way as to minimize correlation between their concentrations (*viz.* that the coefficients of correlation between pairs of concentrations would be less than 0.75).

2.2. Recording of NIR spectra

NIR spectra were recorded in the transmittance mode on a Model 5000 spectrophotometer from FOSS NIRSystems (Silver Spring, MD) equipped with a rapid content analyser (RCA) from the same manufacturer. Each spectrum was recorded by placing ca. 5 g of gel in a quartz cell and the gold reflector on top of the cell; the resulting light path length was 1 mm. A blank spectrum was previously obtained under identical conditions in the absence of sample. Each spectrum, which was obtained by performing 32 scans at 2 nm intervals over the wavelength range 1100–2498 nm, was the average of three

successive measurements between which the gel was turned over.

2.3. HPLC reference method

The determination of the active principle (ketoprofen) and preservatives (methyl and propyl *p*-hydroxybenzoate) using the reference method was performed with an Agilent HP 1100 Series HPLC chromatograph equipped with a Spherisorb S5 ODS2 column and a PDA-UV detector ($\lambda = 258$ nm), using a mobile phase consisting of a 60:40 (v/v) mixture of buffer at pH 3 and acetonitrile. The peak area for each analyte was interpolated into a calibration plot obtained from standard solutions of appropriate concentrations for each analyte.

2.4. Processing of data

Calibration models were constructed by using the software Unscrambler V.9.1. Derivative spectra were obtained by using the Savitzky-Golay algorithm with 11-point moving window and fitting to a second-order polynomial. Partial least-squares (PLS) calibration models were constructed by cross-validation, using the *leave-one-out* method. The goodness of the models was assessed in terms of their standard relative errors (*viz.* the *relative standard error of calibration*, % RSEC, and *relative standard error of prediction*, % RSEP). The number of PLS factors used to construct the models was that resulting in the lowest error of prediction for an external set of samples.

3. Results and discussion

For a calibration model to be accurate, precise and robust enough, the calibration set used to construct it should include every potential source of physical and chemical variability in the samples to be analysed. Chemical (composition) variability is usually incorporated into the calibration set by including samples spanning a range of concentrations below and above the nominal value for the analyte allowing any sample from a production batch to be determined reasonably accurately.

Fig. 1 shows the NIR spectra for the pharmaceutical preparation, its active principles and its preservatives. The spectrum for the preparation exhibits two broad, strong bands at 1470 and 1920 nm corresponding to the first overtone of the O–H bond and the combination band of water. Water is the major excipient (it accounts for 54% (m/m) of the preparation) and also that exhibiting the greatest absorbance; consequently, its characteristic bands predominate in the spectrum for the pharmaceutical. The spectrum additionally contains some bands in the 2300 nm region corresponding to ethanol (42% (m/m)). The most salient bands for the active principle and preservatives (Fig. 1, top) are in the regions 1500–1800 and 2000–2498 nm; however, their low concentrations are not apparent from the sample spectrum.

Commercially available samples invariably span a limited API content range (usually $\pm 5\%$ around the nominal value) which is too narrow to construct effective calibration models. The low contents in the preservatives and their high spectral similarity (see Table 1) required carefully designing the set of laboratory samples to be used for calibration, and

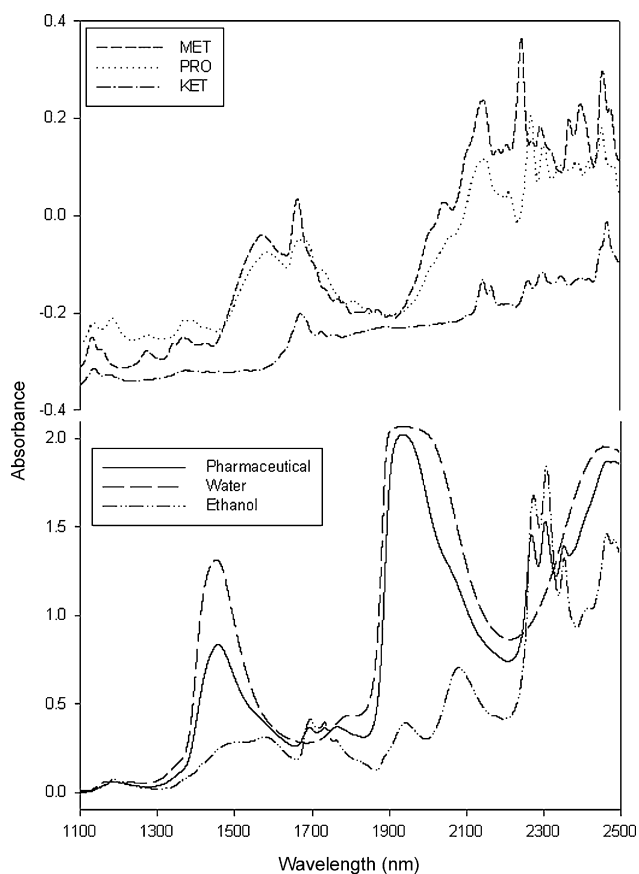


Fig. 1 – NIR spectra of the active, preservatives, main components and pharmaceutical preparation.

appropriate selection of the spectral treatment and wavelength range to be used. Our laboratory samples expanded the content range $\pm 50\%$ around the nominal values; also, correlation between the concentrations of analyte pairs was

Table 1 – Spectral correlation between analytes to quantify

	KET	MET	PRO
KET	1		
MET	0.872	1	
PRO	0.887	0.963	1

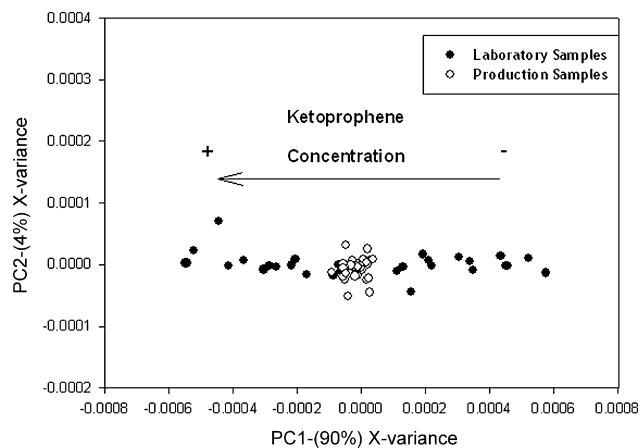


Fig. 2 – PCA score plot from second-derivative NIR spectra of synthetic and production samples.

minimized by using a D-optimal design. A PC2 versus PC scores plot obtained by Principal Component Analysis (PCA) of the second-derivative spectra recorded over the whole wavelength range (1100–2498 nm) from production and laboratory samples (Fig. 2) revealed the ketoprofen concentration to be the greatest source of variability in the spectra and all other sources to be comparatively insubstantial. The plot of Fig. 2 clearly reveals that it was completely unnecessary to include any production samples in the calibration curve in order to obtain an accurate calibration model for ketoprofen (Blanco and Alcalá, 2005).

Table 2 – The models for determination of the three active principles

	KET	MET	PRO
Spectral mode	Second-derivative	Second-derivative	Second-derivative
Spectral range (nm)	1638–1664 2118–2204	1624–1752 2100–2260	1832–2490
No PLS factors	2	5	6
Calibration set			
No. of samples			
Laboratory	19	18	19
Concentration range (mg/g)	12.03–35.99	0.443–1.219	0.135–0.267
RMSEC	0.389	0.032	0.008
Prediction set			
No. of samples			
Laboratory	14	8	10
Production	14	11	13
Concentration range (mg/g)	14.50–36.57	0.358–1.066	0.156–0.262
RMSEP	0.569	0.062	0.035

Table 2 shows the figures of merit for the models used to determine the three analytes. The second-derivative spectral treatment was that providing the best results in all instances. Based on the jack-knifing criterion (Martens and Martens, 2000), we used the spectral ranges listed in Table 2. A comparison of such ranges with the spectra for the pure analytes in Fig. 1 reveals that the selection coincided with the regions of appearance of the most salient bands for the three analytes. The number of PLS factors needed to construct the models increased with decreasing concentration of analyte and Table 2 shows the figures of merit of the three models; as can be seen, they are quite acceptable.

However, determining the contents in ketoprofen and the two preservatives of air-exposed samples revealed that the results were subject to errors of prediction that increased with increasing exposure time. This can be ascribed to ethanol gradually evaporating and altering the sample composition upon exposure to the environment. The API predicted values invariably exceeded the nominal value; by contrast, the preservative predicted contents were slightly lower than the respective nominal values. An ANOVA of the results for the API exposed a significant effect of the time samples were exposed to the air. However, the ANOVA for the preservatives revealed the absence of a similar effect on them.

In order to suppress this source of error in the determination of the API and preservatives, we used a new D-optimal design to prepare 10 new samples consisting solely of water, ethanol and API (i.e. none of the other ingredients) that expanded the ethanol concentration range in order to incorporate the chemical variability of ethanol into the model. The new samples were added to the three calibration sets in order to correct the effect of the evaporation of ethanol. To this end, we constructed new calibration models for the three components. Their figures of merit are shown in Table 3. The new models were slightly more complex than the previous ones and required a greater number of factors for the API and MET; also, their predictive abilities for the three analytes were slightly lower (RMSEPs were slightly greater) as a result

Table 4 – Identification of the pharmaceutical and its components

Compound	Correlation coefficient
Pharmaceutical	0.99
Ethanol	0.68
PRO	0.50
Water	0.35
Jelly agent	−0.33
MET	−0.11
KET	0.10
Threshold 0.95 (positive identification: id results > 0.95).	

of their increased complexity. The predictive ability of the new models was assessed by determining ketoprofen and the preservatives in samples where ethanol was allowed to evaporate partly by exposure to the air and spectra were recorded at 10 min intervals over a period of 30 min. The spectra thus obtained were quantified by using the new models and the results subjected to ANOVA. The new model for quantifying ketoprofen exhibited constancy in its predicted values with time; also, the ANOVA confirmed the absence of significant differences between the results obtained at different times and hence that the length of sample exposure to the air had no significant effect on the predicted values.

The new models for the preservatives suppressed the previous increase in the error of prediction with increasing exposure time. Also, the ANOVA confirmed not only the absence of an effect of the evaporation time, but also that the predictions for the three pharmaceutical ingredients were more accurate.

3.1. Validation of the proposed method

The proposed NIRS method for the simultaneous determination of the API ketoprofen (KET) and the preservatives methyl *p*-hydroxybenzoate (MET) and propyl *p*-hydroxybenzoate

Table 3 – The new models after correct the evaporation effect for determination of the three active principles

	KET	MET	PRO
Spectral mode	2nd Der	2nd Der	2nd Der
Spectral range (nm)	1638–1708 2118–2234	1620–1746 2100–2256	1832–2490
No. of PLS factors	3	6	6
Calibration set			
No. of samples			
Laboratory	29	22	25
Concentration range (mg/g)	12.03–37.16	0–1.071	0–0.303
RMSEC	0.718	0.069	0.015
Prediction set			
No of samples			
Laboratory	14	16	15
Production	14	14	14
Concentration range (mg/g)	14.50–36.47	0.560–1.019	0.129–0.279
RMSEP	0.797	0.166	0.045

Table 5 – Results obtained in the validation of the models

Parameter	Component	KET	MET	PRO
Linearity	n	13	10	10
	Concentration range (mg/g)	14.50–36.47	0.560–1.219	0.151–0.277
	Intercept	-1.38 ± 1.79	0.05 ± 0.211	0.02 ± 0.06
	Slope	1.06 ± 0.08	1.01 ± 0.22	0.91 ± 0.28
	R	0.998	0.959	0.943
Accuracy	n	28	30	29
	Avg difference (mg/g)	-0.23	0.004	0.005
	S.D.	0.776	0.111	0.036
	t experimental	1.127	0.221	0.794
	t tables	2.052	2.045	2.048
Repeatability	Mean NIR (mg/g)	25.40	0.722	0.180
	CV (%)	1.89	8.01	9.67
Intermediate Precision	Day			
	F experimental	0.02	3.24	2.27
	F tables	9.55	9.55	9.55
	Analyst			
	F experimental	2.32	0.329	1.82
	F tables	7.70	7.70	7.70
	CV (%)	1.23	3.32	6.03
Robustness	n	30	30	30
	t experimental	1.816	1.754	1.804
	t tables	2.045	2.045	2.045

(PRO) was validated in accordance with the ICH (Note for guidance, 2003) and EMEA guidelines (ICH, 1994). To this end, we assessed its selectivity, linearity, accuracy, precision (as repeatability and intermediate precision) and robustness.

At difference to other analytical techniques, selectivity is applied to a NIRS methodology in a different way; its requirements have been studied and solved in the NIRS specific guidelines. The selectivity of a NIRS method is established by the identification of the pharmaceutical preparation and not to the presence of the analyte.

The selectivity of the method was established by identifying the gel prior to quantifying the analytes. The gel was identified by using a spectral library containing 22 spectra for the preparations that were obtained from seven production batches. The discriminating criterion used for identification was the correlation coefficient for the second-derivative spectra over the entire wavelength range (1100–2498 nm). A threshold of 0.95 allows the pharmaceutical preparation to be accurately identified and discriminated from the other products.

The selectivity of the library was assessed by using it to identify external samples; all productions samples thus processed were accurately identified as well. Table 4 shows the identification results. The analysis routine used precluded determining a negatively identified sample.

All other properties were validated in accordance with the above-mentioned guidelines (see Table 5). The absence of treatment of the sample lead to validate the robustness of the proposed method from the results obtained for samples that were analysed over a period of 1 year. The application of a t-test at 95% confidence level (Miller and Miller, 2005) on the differences between the NIRS and HPLC values (paired test significance) revealed that the results of the two methods do

not show significant differences (Blanco et al., 2004) between them.

4. Conclusions

Using NIRS to analyse samples with a highly absorbing matrix (e.g. hydroalcoholic gels) containing ingredients at low concentrations requires exercising extreme care in constructing the calibration models to be used. The high spectral correlation between the analytes – which precludes the construction of simple, robust models – can be corrected by preparing samples with a low correlation between concentrations; also, the low concentrations of some analytes, which reduce their contributions to the sample spectrum, require careful selection of the wavelengths and spectral mode to be used to construct the models. This paper testifies to the suitability of NIRS for the analytical control of such kind of samples. In combination with appropriate models, the NIRS technique allows the accurate analysis of pharmaceutical preparations with a complex matrix and low contents in its ingredients.

Acknowledgements

The authors are grateful to Spain's MCyT for funding this research within the framework of Project CTQ2006-12923. Also, they wish to thank Laboratorios Menarini, S.A. (Badalona, Spain) for kindly supplying the studied samples.

REFERENCES

Blanco, M., Alcalá, M., 2005. Use of near-infrared spectroscopy for off-line measurements in the pharmaceutical industry. In:

- Bakeev, A.K. (Ed.), *Process Analytical Technology*, 1st ed. Blackwell Publishing, Oxford, pp. 375-376 (Chapter 10).
- Blanco, M., Alcalá, M., 2006. Simultaneous quantitation of five active principles in a pharmaceutical preparation: development and validation of a near infrared spectroscopic method. *Eur. J. Pharm. Sci.* 27 (2-3), 280-286.
- Blanco, M., Romero, M.A., Alcalá, M., 2004. Strategies for constructing the calibration set for a near infrared spectroscopic quantitation method. *Talanta* 64, 597-602, 2004.
- Colón, J., Peroza, C., Caraballo, W., Conde, C., Li, T., Morris, K.R., Romañach, R.J., 2005. On line non-destructive determination of drug content in moving tablets using near infrared spectroscopy. *Proc. Anal. Technol.* 2 (5), 8-15.
- ICH Guidelines: Q2A Text on Validation of Analytical Procedures, 1994. Q2B Validation of Analytical Procedures: Methodology 1996. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceutical for Human Use.
- Jing, Q., Shen, Y., Tang, Y., Ren, F., Yu, X., Hou, Z., 2006. Determination of nelfinavir mesylate as bulk drug and in pharmaceutical dosage form by stability indicating HPLC. *J. Pharm. Biomed. Anal.* 41, 1065-1069.
- 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.
- Miller, J.N., Miller, J.C., 2005. Significance test. Comparison of an experimental mean with a known value. In: *Statistics and Chemometrics for Analytical Chemistry*, 5th ed. PEARSON Prentice Hall.
- <http://www.emea.eu.int/pdfs/human/qwp/330901en.pdf>, 2003. Note for guidance. The European Agency for the Evaluation of Medicinal Products.
- Quaglia, M.G., Donati, E., Fanali, S., Bossu, E., Montinaro, A., Buiarelli, F., 2005. Analysis of diltiazem and its related substances by HPLC and HPLC/MS. *J. Pharm. Biomed. Anal.* 37 (4), 695-701.

6.3 API DETERMINATION BY NIR SPECTROSCOPY ACROSS PHARMACEUTICAL PRODUCTION PROCESS

M. Blanco, M. Bautista, M. Alcalà
AAPS PharmSciTech. 2008;9 (4):1130-5.

Research Article

API Determination by NIR Spectroscopy Across Pharmaceutical Production Process

M. Blanco,^{1,3} M. Bautista,¹ and M. Alcalá²

Received 5 May 2008; accepted 18 September 2008; published online 14 November 2008

Abstract. The purpose of this research was to demonstrate the ability of reflectance near-infrared (NIR) spectroscopy for quantitative analysis of an active ingredient in different production steps of a solid formulation. The drug is quantified at two different steps of a pharmaceutical process: after granulation and after tablet coating. Calibration samples were prepared by mixing pure drug, excipients, and batch samples (75–120 mg/g active ingredient) using a simple methodology that can be easily carried out in a laboratory. Partial least squares calibration models were calculated in second-derivative mode using the wavelength range 1,134–1,798 nm. The error of prediction for granulated samples was 1.01% and 1.63% for tablets. The results prove that NIR spectroscopy is a good alternative to other, more time-consuming means of analysis for pharmaceutical process monitoring.

KEY WORDS: API determination; NIR spectroscopy; PAT; PLS calibration; tablet content uniformity.

INTRODUCTION

Quality control constitutes an essential element of pharmaceutical production processes by virtue of the high safety levels demanded from commercially available formulations. Assuring quality in such formulations entails subjecting the end products to a wide variety of tests; this has an adverse impact on production expeditiousness and commercial competitiveness. In response to this problem, the FDA PAT Guideline (1) established a new working philosophy aimed at replacing end-product controls with monitoring of the whole production process in order to identify any weaknesses with a view to facilitating the correction and assuring quality in the target product by acting throughout the process. One major objective of this initiative was to improve the quality of end products while increasing the efficiency of the production processes; in this way, quality would be built across the process. The best process knowledge, the higher probability to improve the quality of the product.

Fulfilling this objective requires a deep knowledge of the effect of some variables on the process and such knowledge can be acquired by determining the critical parameters and monitoring the changes in order to correct any undesirable deviations.

Implementing the new working philosophy requires the development and use of simple, rapid analytical techniques to expeditiously acquire the information needed (2). Spectro-

scopic techniques are especially amenable to this use. This is especially so with near-infrared spectroscopy (NIRS). In fact, the NIRS technique has been widely used with success in this context, as reflected by the large number of reported applications in the pharmaceutical industry (3–6), where its flexibility has enabled the obtainment of analytical information for untreated samples of any type with minimal human intervention.

Quantifying the active principle (API) in different steps of a production process and its end product usually helps assess the status of the process and confirm that a given production batch is fit for release. API quantitation usually relies on high performance liquid chromatography (7,8), which uses time-consuming protocols. This has led to its increasing replacement with NIR spectroscopy in recent years, where this technique has proved a useful tool for the simultaneous determination of physical and chemical properties of samples in a clean, expeditious, and non-destructive manner in various steps of the production process (9–11).

Wet granulation is a basic operation intended to improve the flowability of pharmaceutical mixtures because of an increase in the mean particle size. In addition, the granulation process is a common step where the end product is released in form of tablets or granulated mixtures.

Determinations by NIR spectroscopy require the use of multivariate calibration models encompassing all potential sources of variability during routine analysis. Because the end product invariably spans too narrow a concentration range, the models must be constructed from additional, laboratory samples of variable enough concentrations to ensure the required accuracy, precision, and robustness (11). One easy way of expanding the API content range is by using underdosed and overdosed samples prepared by adding small amounts of the excipients and API. Scheiwe *et al.* (12)

¹Departament de Química, Unitat de Química Analítica, Universitat Autònoma de Barcelona, Bellaterra, 08193, Barcelona, Spain.

²Department of Chemistry, Faculty of Arts and Sciences, The University of Puerto Rico, Mayagüez, 00681, Puerto Rico.

³To whom correspondence should be addressed. (e-mail: marcel.blanco@uab.es)

determined the API content in tablets by using Transmission mode. This work presented a successful strategy of calibration and routine API content monitoring; however, the experimental approach involved a great effort of sample preparation in a pilot plant installation. The present work shows an easy and novel strategy to prepare the calibration samples in the laboratory for the development of calibration models to analyze intact tablets by NIR spectroscopy. This strategy simplifies the sample preparation if different calibration models need to be calculated for individual manufacturing step. Basically, a few number of each manufacturing step is included in the calibration sample set to include the process variability in the multivariate model.

Altering the amounts of API and excipients in the production process to construct a calibration set with a wide range of API represents a high cost for the company. In the present work, production tablets at nominal contents were milled into powder samples. Once these samples were uniformly milled, API and excipients were added obtaining a calibration set including the chemical variability required to correctly predict production samples. This strategy supposes a great diminution in time and cost.

In this work, two PLS calibration models were redeveloped in order to quantify the API in the pharmaceutical product after two different steps of the production process, namely: wet granulation of the powdered components and tableting and coating this granulate. The first calibration model provides significant information about the conformity parameters of the obtained granulated mixture to be used in the next steps of the process. The second calibration model allows monitoring the API content from the obtained coated tablets.

Thus, the calibration models obtained were applied to various steps of the production process and the two ensuing quantitation methods were subsequently validated in accordance with the ICH and EMEA guidelines (13,14).

MATERIALS AND METHODS

Pharmaceutical Formulation

The formulation studied consisted of tablets containing 135 mg/g dexketoprofen trometamol as API; PH101 microcrystalline cellulose (500 mg/g) as major excipient; and maize starch, carboxymethyl starch, and glycerol distearate as minor excipients. The tablets were coated with a thin film by spraying a solution consisting of OPADRY lacquer (85%) and propylene glycol (15%) diluted in purified water, and finally dried.

Although dexketoprofen (DKP) was present as its trometamol salt, the reference values were expressed as DKP (1 g of DKP in acid form is equivalent to 1.476 g of DKP trometamol). The nominal content in DKP of the production samples was 93 mg/g.

All studied samples were supplied by Laboratorios Menarini, S.A. (Badalona, Spain).

Production of the Pharmaceutical

The production process of the tablets involves several steps including joint granulation of the API and some

excipients, addition of the other ingredients, blending, pressing, and coating. The tablets are cylindrical in shape and weigh 270 mg on average.

In the first granulation step, a mixture of DKP, cellulose, and starch is granulated in a fluid bed reactor that is supplied with a binder solution and then dried; the average duration of this step is 270 min. The resulting granulate is then homogeneously blended with the other excipients (carboxymethyl starch and glycerol distearate) in the second step. Next, the granulated product is pressed and homogenized to obtain cores (non-coated tablets). In the fourth, final step, the tablets are coated with a film of lacquer.

Two calibration models to quantify the API after blending of the granulate and excipients, and also after coating of the tablets (*i.e.*, after the second and fourth steps), were constructed.

Samples

Calibration models were constructed by using sets of samples obtained from different batches depending on the particular step of the production process.

Production Samples

The production samples were coated tablets from 20 different production batches obtained over a period of 6 months. The calibration set for tablets was constructed from intact coated tablets. On the other hand, the model for the granules (end of the second step) was constructed from tablets that were previously milled and underdosed or overdosed to expand the concentration range.

All samples were additionally analyzed by using a UV-Vis reference method developed and validated by Laboratorios Menarini, S.A.

Underdosed and Overdosed Samples

Commercial samples invariably span a narrow API concentration range (around $\pm 5\%$ of the label claim), so these samples are unsuitable for constructing calibration models. Therefore, the API content range must be expanded by using additional laboratory samples prepared by underdosing and overdosing milled production tablets.

The samples used to construct the two calibration models were milled tablets from production batches and doped (underdosed and overdosed) samples spanning the DKP concentration range 75–120 mg/g.

Regardless of introducing physical variability in the calibration set (particle size), milling production tablets in the laboratory allows expanding the API range. One alternative strategy could be realized in the granulation process in the laboratory, but the time consumption is higher and, moreover, this manual granulation can provide some variability not included in the production process.

Overdosed samples were prepared from powdered production tablets that were supplied with accurately weighed amounts of API. *Underdosed samples* were also prepared from powdered production tablets which, however, were supplied with known amounts of a mixture of the excipients in the same proportions as in the pharmaceutical preparation.

Both types of samples were mixed in a Turbula shaker until the NIR spectra exhibited no appreciable changes between consecutive measurements.

Recording of NIR Spectra

Spectra were recorded on a model 5000 spectrophotometer from Foss NIRSystems (Silver Spring, MD, USA) equipped with a rapid content analyzer module and governed via the software Vision v. 2.51. A ceramic plate was used to obtain a reference spectrum prior to measuring each sample. All spectra were the averages of 32 scans performed at 2-min intervals over the range 1,100–2,498 nm.

Aliquots of production and doped granulated samples were placed in a quartz cell and the reflectance spectra were recorded in triplicate with turnover between recordings. The resulting average spectra were used to construct and validate the models, and also to quantify the API in the pharmaceutical preparation.

The spectra for coated production tablets were obtained for both sides by placing each tablet on the quartz window of the recording module. The spectra for the two sides were averaged for subsequent processing.

Processing of NIR Spectra

PLS1 models were based on various sample sets to determine the API in two steps of the production process. To this end, the software Unscrambler v. 9.2 from CAMO Process AS (Oslo, Norway) has been used.

The spectral pretreatments used to construct the calibration models included standard normal variate and first and second derivatives. Derivative spectra were obtained by using the Savitzky–Golay algorithm with an 11-point moving window and a second order polynomial. The models were obtained by cross-validation and the optimum number of factors was taken to be that resulting in the minimum PRESS value (Eq. 1) and the quality of the resulting models was assessed in terms of the relative standard errors of calibration (% RSEC) and prediction (% RSEP) (Eq. 2):

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

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

where \hat{y}_i and y_i correspond to the drug concentration of each sample predicted by the NIR method and reference method, respectively.

UV Reference Method

The API (dexketoprofen) contents of the production samples were determined by ultrasonically treating an amount of ca. 0.25 g of milled sample in 90 ml of Milli-Q water at 37°C for 15 min. Then, an aliquot of 2 ml was diluted to 50 ml of a 3:1

MeOH/H₂O mixture and its UV spectrum recorded. The dexketoprofen content was obtained by multiple linear regression of the 210–400 nm region of the spectrum, using that for pure API in MeOH/H₂O as reference.

RESULTS AND DISCUSSION

We aimed at establishing simple calibration models with an adequate predictive ability for use in various steps of the production process with a view to facilitating the straightforward, expeditious determination of the dexketoprofen content of the product. To this end, two calibration models were constructed to be used after the granulation step (model 1) and for quality control of the end product (coated tablets) prior to its release (model 2). Both models were constructed from samples in two different forms, namely: milled coated tablets and intact coated tablets. Following validation, each calibration model was used to determine the API content in the specific step of the production process in order to check its appropriateness and allow the process to continue. Obtaining an accurate, precise, and robust calibration model requires using a calibration set consisting of samples encompassing all potential sources of physical and chemical variability in the samples to be subsequently determined.

Figure 1 shows the absorbance spectra for the API and the two types of samples studied (milled tablets, non-coated and coated tablets). As can be seen, the three spectra exhibited an identical profile (*i.e.*, no difference in chemical composition between samples was apparent); however, the spectra were shifted by effect of differences in physical properties between samples. Thus, the tablets exhibited higher absorbance values by effect of the pressing; the higher the pressure used was, the higher was the absorbance (3). Because the sole difference between the two types of tablets was the presence of the lacquer coating—which was in a very low proportion and absorbed very slightly—in one, the spectra were essentially identical.

Analysis of Granulated Samples

Model 1 was constructed from milled coated tablets in addition to milled tablets underdosed and overdosed

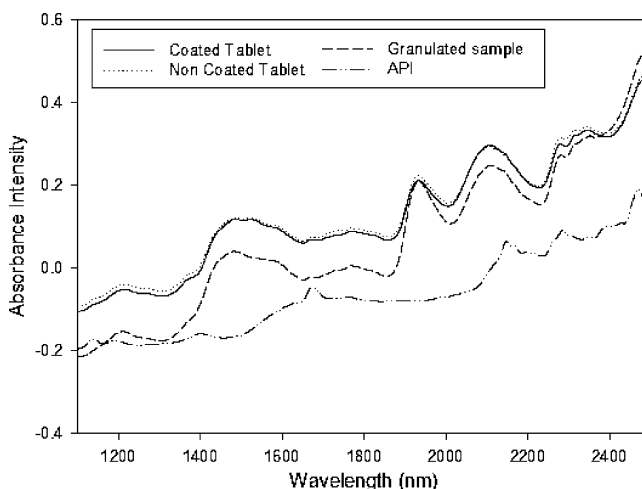


Fig. 1. NIR spectra for coated and non-coated tables, granulated samples and the pure API

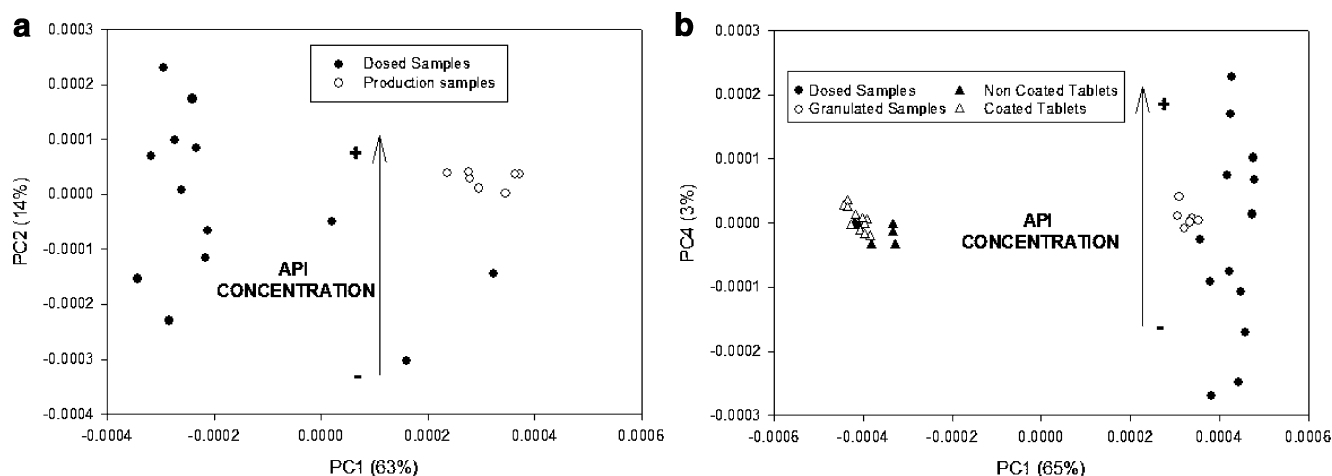


Fig. 2. **a** PCA scatter plot from absorbance spectra over 1,134–1,798 nm for dosed and production samples. **b** PCA scatter plot from absorbance spectra over 1,134–1,798 nm for coated and non-coated tablets, granulated samples, and dosed samples

intended to expand the API content range in order to encompass the whole chemical variability of the production process. These samples were used in combination with various spectral ranges and treatments. Figure 2a shows the scatter plot of calibration samples (production samples with nominal content of API and over/subdosed samples covering the API range). PC2 was the factor responsible to the API variation. Production samples are grouped while dosed samples are expanded through the PC2. Table Ia shows the most salient parameters of the model samples and the predictions obtained for granulated production samples. Second-derivative spectra were those providing the best results. The model was constructed from 20 samples of which 13 were of the dosed type and seven contained the nominal API level. These three PLS factors provided acceptable calibration statistics (99.1% Y-explained variance) and predictive ability (RSEP=1.01%) with granulated production samples. As can be seen from Table I, the model exhibited a good predictive ability with granulated samples. Once this model was capable of correctly predicting samples at the end of the granulation process, coated and non-coated tablets were evaluated with this model to test the chance of using the same model at the end of the process.

The model for granulated samples was used to analyze tablets, both cores and coated. As can be seen from Fig. 3a, the ensuing predictions were invariably overestimated and subject to an RSEP greater than 5%; this suggests that differences in compactness between samples resulted in substantial errors, so the model was unsuitable for tablets.

Analysis of Tablets

The NIR spectra of granulated samples used in the previous model and the spectra of coated and non-coated tablets were subjected to principal component analysis (PCA). Figure 2b shows the scatter plot for a PCA of second-derivative spectra obtained with the Savitzky–Golay algorithm. As can be seen, the first factor (PC1) essentially contained the physical differences between sample types exposed by Fig. 1. Based on the results of Fig. 2, PC4 was the factor that more clearly showed the chemical variability of

the samples; thus, while the coated tablets—which contained virtually the same amount of API—lay in the center of the PC4 axis, the milled samples—which included underdosed and overdosed samples—lay in different positions along the axis depending on the particular API content.

Constructing a model capable of accurately predicting the API contents of intact tablets required using samples spanning a wide content range. The inability to alter coated tablets to this end led us to use the set of samples previously employed to expand the range for model 1. The model was constructed from 13 granulated samples spanning the API content and five intact coated tablets containing the nominal API level. This strategy allowed including both chemical variability and physical variability—a result of pressing the granulated samples, mainly—to be incorporated by using coated production tablets. Although the incorporation of coated tablets into the model introduced chemical variability in the coating due to the small amount of film coating on the tablets, its effect on the resulting spectra was minimal (see

Table I. Figures of Merit of the Models for Tablets and Granulated Samples

	Calibration model	
	(a) Granulate	(b) Tablets
Spectral pretreatment	2nd der. S–G	2nd der. S–G
Spectral range (nm)	1,134–1,798	1,134–1,798
No. of PLS factors	3	3
Y-explained variance %	99.12	99.31
Calibration set		
Dosed and milled samples	20	13
Tablets	–	5
Concentration range (mg/g)	75.6–120.6	75.6–120.6
RSEC	1.04	0.97
Prediction set		
Granulated samples	20	–
RSEP	1.01	–
Tablets	28	28
RSEP	5.03	1.63

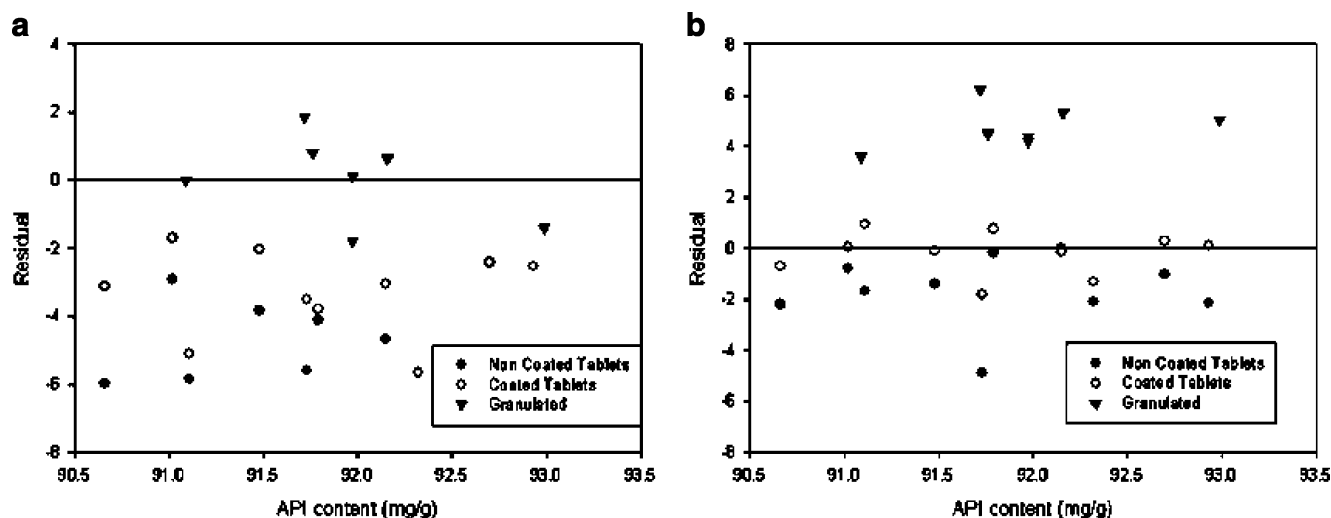


Fig. 3. Residuals obtained in the prediction of tablets and granulated samples with models 1 a and 2 b

Fig. 1). In this figure, differences in the spectra of coated and non-coated tablets cannot be observed.

Table Ib shows the figures of merit of the PLS1 model as used to predict coated tablets at the end of the production

Table II. Validation Parameters for the API Quantitation with the Models for Tablets and Granulated Samples

		Model	
		Granulate	Tablets
Selectivity	Compound	Correlation coefficient	
	Enantyum granulated	>0.99	–
	Coated tablets	–	1
	Uncoated tablets	–	0.99
	Maize starch	0.75	0.76
	Microcrystalline cellulose	0.90	0.86
	Carboxymethyl starch	0.79	0.70
	Dexketoprofen	0.31	0.30
	Glycerol palmitoestearate	0.24	0.33
	Parameter	Results	
Linearity	<i>n</i>	20	8
	Concentration range (mg/g)	75.6–120.6	74.1 – 112.5
	Intercept	–0.02±4.54	3.9±10.6
	Slope	1.00±0.05	0.96±0.12
	<i>R</i>	0.995	0.993
Accuracy	<i>n</i>	10	10
	Avg. difference (mg/g)	–0.33	0.17
	SD	1.87	1.00
	<i>t</i> exp	1.11	0.54
	<i>t</i> crit	2.23	2.26
Repeatability	Replicates	6	6
	Mean NIR (mg/g)	94.64	91.22
	CV (%)	0.39	0.21
Intermediate	Days (3)	F exp 13.7	F exp 0.66
		F tab 19.00	F tab 19.00
Precision	Analysts (2)	F exp 0.07	F exp 0.09
		F tab 18.51	F tab 18.51
		CV (%)	0.72
Robustness	<i>n</i>	20	14
	<i>t</i> exp	0.56	0.41
	<i>t</i> crit	2.09	2.16

process (model 2). The model was constructed by using various spectral ranges and treatments, and also different types of samples. The best choice was that based on the same parameters as the previous model, namely: three PLS factors, which sufficed to accurately predict production tablets.

The predictions for coated tablets had an RSEP of 1.63%, which is substantially lower than that provided by model 1; also, the residuals were randomly distributed around 0, so the model can be deemed accurate for predicting samples at the end of the process.

Using model 2 to analyze non-coated tablets (cores) provided predictions with an RSEP of 2.30%. The predictions were systematically overestimated (see Fig. 3b) for all batches, which can be ascribed to the absence of the lacquer coating.

Finally, model 2 was used to predict granulated samples, with which it gave an RSEP of 3.43%. Contrary to non-coated samples, all samples were underestimated (Fig. 3b).

Validation of the NIR Method for Determining the API

The two NIR models were validated prior to use in routine analyses of the intermediate granulate obtained in the production process and also of individual coated tablets—

Table III. API Concentration in Coated Tablets as Determined with Model 2

Reference value	Prediction (mg/g)	Residual	Residual %
91.480 mg/g	92.112	–0.632	–0.69
	91.217	0.263	0.29
	92.154	–0.674	–0.74
	91.833	–0.353	–0.39
	89.906	1.574	1.72
	89.809	1.671	1.83
	89.897	1.583	1.73
	93.281	–1.801	–1.97
	92.129	–0.649	–0.71
	93.244	–1.764	–1.93
Mean value	91.558	–0.078	–0.09

which were assessed for content uniformity. To this end, the validation guidelines of the *International Conference on Harmonisation* (ICH) and the *European Agency for the Evaluation of Medicinal products* (EMA) were followed to assess the selectivity, linearity, accuracy, precision (as repeatability and intermediate precision), and robustness of the proposed method.

The selectivity was established by identifying the sample prior to quantifying the analyte. Identification was based on a previously built library of second-derivative spectra for the pharmaceutical and its individual ingredients recorded over the whole spectra range (1,100–2,498 nm); identification was deemed positive if the correlation coefficient exceeded the threshold (0.98). All production samples studied were positively identified and all pure components in the formulation were accurately discriminated, with correlation coefficients lower than the threshold.

By contrast, the model for tablets required distinguishing between coated and non-coated tablets; this was done by using a sub-library of second-derivative spectra obtained over the spectral range 1,100–2,188 nm. The discriminating criterion used was the Euclidian distance in the space of three PCs, with a threshold of 0.46. This sub-library allowed coated tablets to be accurately distinguished from non-coated tablets (cores).

Table II shows the most salient results for the other validation parameters of the granulated product and tablets as determined in accordance with the ICH and EMA guidelines for NIR methods (12). The proposed NIR method met all requirements in both guidelines, which testifies to its usefulness as a routine analytical method for the pharmaceutical industry.

Once model 2 was validated, the determination of API concentration in coated tablets was assessed. The US Pharmacopoeia recommends individual analyses of ten tablets from the same batch and using the average value for the ten determinations in order to determine content uniformity in production batches.

While residuals of up to 15% in individual analyses are deemed acceptable, the average residual should not exceed 5% of the nominal value. Table III shows the results of API concentration in ten coated tablets in the same batch with the validated model 2.

CONCLUSIONS

The use of NIR reflectance spectroscopy allows the API content of a pharmaceutical preparation to be controlled throughout the production process and batches to be checked for acceptability in this respect prior to commercial release. The two models were constructed by using a low number of samples prepared in the laboratory and provided accurate predictions (errors lower than 1.7%). The methodology used to construct the models can be extended to more complex processes involving a greater number of steps with little additional effort.

ACKNOWLEDGEMENTS

The authors are grateful to Spain's MCyT for funding this research within the framework of Project CTQ2006-12923. Also, we wish to thank Laboratorios Menarini, S.A. (Badalona, Spain) for kindly supplying the studied samples.

M. Alcalá thanks the National Science Foundation Engineering Research Center for Structured Organic Particulate Systems for the postdoctoral grant funding.

REFERENCES

1. Guidance for Industry. PAT—A framework for innovative pharmaceutical development, manufacturing, and quality Assurance. US Department of Health and Human Services, Food and Drug Administration, Center of Drug Evaluation (CDER), and Research, Center for Veterinary Medicine (CVM), Office of Regulatory Affairs (ORA), 2004.
2. S. Kueppers, and M. Haider. Process analytical chemistry—future trends in industry. *Anal. Bioanal. Chem.* **376**:313–315 (2003).
3. P. Chalus, Y. Roggo, S. Walter, and M. Ulmschneider. Near-infrared determination of active substance content in intact low-dosage tablets. *Talanta*. **66**:1294–1302 (2005).
4. M. Blanco, and M. Alcalá. Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy A contribution to process analytical technologies. *Anal. Chim. Acta.* **557**:353–359 (2006).
5. M. Blanco, M. Alcalá, J. M. González, and E. Torras. 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**:2137–2144 (2006).
6. C. Peroza, M. A. Santos, and R. J. Romañach. Quantitation of drug content in a low dosage formulation by transmission near infrared spectroscopy. *AAPS Pharm. Sci. Tech.* **7**(1) (2006), Article 29.
7. C. Sajeev, P. R. Jadhav, D. RaviShankar, and R. N. Saha. Determination of flurbiprofen in pharmaceutical formulations by UV spectrophotometry and liquid chromatography. *Anal. Chim. Acta.* **463**:207–217 (2002).
8. A. Abourashed, and A. Khan. High-performance liquid chromatography determination of hydrastine and berberine in dietary supplements containing goldenseal. *J. Pharm. Sci.* **90**:7817–822 (2001).
9. M. Donoso, D. O. Kildsig, and E. S. Ghaly. Prediction of tablet hardness and porosity using near-infrared diffuse reflectance spectroscopy as a nondestructive method. *Pharm. Dev. Tech.* **8**:357–366 (2003).
10. A. C. Peinado, F. van den Berg, M. Blanco, and R. Bro. Temperature-induced variation for NIR tensor-based calibration. *Chemom. Intel. Lab. Sys.* **83**:75–82 (2006).
11. M. Blanco, and M. Alcalá. Use of near-infrared spectroscopy for off-line measurements in the pharmaceutical industry. In K. Bakeev (ed.), *Process Analytical Technology*, Blackwell, Oxford, 2005, pp. 375–376.
12. M. W. Scheiwe, D. Schilling, and P. Aebi. Near Infrared spectroscopy analysis of intact diclofenac coated tablets in transmission. *Pharmazeutische Industrie.* **61**:179–183 (1999).
13. ICH Q2B: Validation of Analytical procedures: Methodology, Consensus Guideline, International Conference on Harmonisation (ICH), 1998. <http://www.bcg-usa.com/docs/ich/ICHQ2B.pdf>
14. Note for guidance on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations, 2003. <http://www.fda.gov/Cber/gdlns/cmprprot.pdf>

6.4 ON-LINE MONITORING OF GRANULATION PROCESS BY NIR SPECTROSCOPY

M. Alcalà, M. Blanco, M. Bautista. J.M. González
Accepted in Journal of Pharmaceutical Sciences.

On-Line Monitoring of A Granulation Process By NIR Spectroscopy

MANEL ALCALÀ,¹ MARCELO BLANCO,¹ MANEL BAUTISTA,¹ JOSEP M. GONZÁLEZ²

¹Departament de Química, Unitat de Química Analítica, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain

²Laboratorios Menarini, c/Alfons XII, 587, E-08918 Badalona, Barcelona, Spain

Received 5 December 2008; revised 17 March 2009; accepted 19 April 2009

Published online ? ? ? ? in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/jps.21818

ABSTRACT: Near infrared (NIR) spectroscopy has been used in a noninvasively mode to develop qualitative and quantitative methods for the monitoring of a wet granulation process. The formulation contained API (10%w/w) and microcrystalline cellulose and maize starch as main excipients. NIR spectra have been acquired through the glass window of the fluidizer in reflectance mode without causing interference to neither the process nor the formulation. The spectral data has been used to develop a qualitative multivariate model based on principal component analysis (PCA). This qualitative model allows the monitoring of different steps during the granulation process only using the spectral data. Also, a quantitative calibration model based on partial least squares (PLS) methodology has been obtained to predict relevant parameters of the process, such as the moisture content, particle size distribution, and bulk density. The methodology for data acquisition, calibration modeling and method application is relatively low-cost and can be easily performed on most of the pharmaceutical sites. Based on the results, the proposed strategy provides excellent results for the monitoring of granulation processes in the pharmaceutical industry. © 2009 Wiley-Liss, Inc. and the American Pharmacists Association *J Pharm Sci* 9999:1–11, 2009

Keywords: NIR spectroscopy; principal component analysis; partial least squares; granulation; monitoring

INTRODUCTION

Granulation is a process routinely used by the pharmaceutical industry to increase particle size while improving the flowing properties of a solid formulation, facilitating blending of its components and increasing the compressibility of the material for the production of tablets. Wet granulation involves adding a binder solution to a powdered solid mixture in order to facilitate particle aggregation. As the amount of binder added increases, small particles agglomerate by

forming liquid junctions. Capillary, viscous and frictional forces make the resulting agglomerates strong enough to increase particle size in the mixture in order to obtain appropriate granules.¹

Designing, monitoring and controlling a wet granulation process requires a sound knowledge of the variables associated to the process and its product. Traditionally, granulation processes have been monitored by using visual and manual procedures, and the resulting granulates characterized in terms of particle size, moisture content, and bulk and tap density. Some reference methods used to determine these properties require using a substantial amount of granulate in many steps of the process (e.g., during addition of the binder or drying of the product). Although they provide acceptable results, these techniques

Correspondence to: Manel Alcalà (Telephone: +34935811712; Fax: +34935812379; E-mail: manel.alcala@uab.cat)

Journal of Pharmaceutical Sciences, Vol. 9999, 1–11 (2009)

© 2009 Wiley-Liss, Inc. and the American Pharmacists Association

are time-consuming and occasionally call for experienced analysts and/or technicians, which preclude their use in the now increasingly expanding process analytical technology (PAT) strategies.

NIR spectroscopy has become a useful tool for the expeditious extraction of quality information about physical and chemical properties of pharmaceutical preparations.² This technique has enabled the development of noninvasive on-line monitoring methods that provide analytical results allowing running processes to be corrected if needed, the time used in each step shortened and quality in the end-product increased.

The development of technological innovations in the pharmaceutical industry, particularly as regards production, is widely supported, especially by FDA and other regulatory bodies.^{3,4} The primary aim of PAT and quality by design (QbD) strategies is increasing existent understanding and knowledge about production processes. However, the pharmaceutical industry can only implement those PAT tools with lower development, implementation, validation and maintenance costs than their expected benefits. In this situation, the development of straightforward analytical methodologies based on near infrared (NIR) spectroscopy for the better understanding of the processes can eventually improve knowledge about the target product, shortening times in each step and obtaining an economic benefit.

A number of studies have examined the batch determination of homogeneity, particle size and its distribution by NIR spectroscopy,^{5,6} using off-line measurements to facilitate implementation of the required analytical tools. Thus, Sekulic et al.⁷ developed a NIR method for the real-time monitoring of homogeneity in a V-blender. Rantanen et al.⁸ conducted an on-line study of granulation in a pilot plant high shear mixer and used chemometric algorithms to monitor changes in particle size and moisture content. Recently,⁹ several models for quantifying moisture and average particle size during granulation of a single component (the excipient lactose) in a pilot plant mixer results have been proposed. However, controlling and optimizing the variables potentially influencing the outcome of a granulation process (e.g., input air temperature, binder flow-rate, sprayer position, component concentrations) can be easily accomplished at the pilot plant scale. Rambali et al.¹⁰ studied changes in grain size on a small, medium and large scale. Other authors^{11–13} have developed effective experimen-

tal designs for determining physical and chemical parameters during granulation with a view to optimizing the properties of the end-product. Frake et al.¹⁴ developed quantitative methods for moisture and median particle size, using invasive NIR probe on pilot plant batches. Green et al.¹⁵ explained issues related to sampling effects on method accuracy. The authors demonstrated that process heterogeneity plays a major role in determination of apparent prediction accuracy. Macias et al.¹⁶ reviewed different methods (in-line and noninvasive) that successfully track particle size enlargement during high-shear wet granulation.

In the present work, various variables (viz. particle size distribution, moisture content and density) have been monitored during a wet granulation process at a production plant. To this end, a portable NIR spectrometer has been used for the on-line spectral acquiring on a noninvasive approach. The acquired spectral data has been used to develop both qualitative and quantitative methods. Principal component analysis (PCA) and partial least squares (PLS) techniques have been used to monitor the granulation process and to quantify physical and chemical properties of the formulation. The results of both methods have been combined to characterize the formulation and obtain a better knowledge of the granulation process. To the best author's knowledge, there are very few publications where NIR methods were developed on real industrial processes as this done, primarily due to concerns about proprietary issues. For this reason, the authors consider this manuscript a useful reference in the pharmaceutical technology field for obtaining both physical (particle size distribution and bulk density) and chemical (moisture) information from the spectral visualization and data modeling.

MATERIALS AND METHODS

Samples

The pharmaceutical formulation studied is a solid manufactured as tablet containing about 100 mg of active principle ingredient per gram of product, PH101 microcrystalline cellulose (500 mg/g) as major excipient, and maize starch, carboxymethyl starch and glycerol distearate—the latter two are added directly to the granulate—as minor excipients. The active ingredient is a nonsteroidal anti-inflammatory drug.

Microcrystalline cellulose and maize starch are fillers in the formulation diluting the drug to the nominal concentration level. Carboxymethyl starch is used as disintegrant, that allow the fast disintegration of each unit dose in dissolution. Glycerol diestearate is a common lubricant that improves the flowability of the bulk on latter production steps after granulation. All samples and ingredients were supplied by Laboratorios Menarini, S.A. (Badalona, Spain), and all tests were conducted in their facilities before tableting.

Granulation Process

Granulation was performed on a GLATT WSG300 fluid bed system (FBS). The industrial granulation process used 260 kg loads of the raw materials [viz. the API and the two major excipients (microcrystalline cellulose and starch)] under the normal operating conditions leading to the obtainment of a product with the desired specifications.

The raw materials were loaded by aspiration at the top of the reactor. An air stream was used to continuously agitate the load in order to ensure homogeneous mixing of the materials before granulation was started. A 27% of the total maize starch was added to a volume of 200 L of distilled water to prepare the binder solution that was added during the granulation process.

The total process time was 255 min and involved three different steps, namely: agitation of the powder mixture using a hot air stream (15 min), addition of the binder solution (140 min) and drying (100 min). In the first step, an air stream (60°C) was used to ensure thorough mixing of the load prior to granulation. Then, the temperature of the air stream was raised to 80°C and the binder solution was added as a spray (1.5 L/min) on the top of the fluidizer. Finally, an air stream (90°C) was used to dry the granulated mixture.

Determination of Reference Parameters

Aliquots of 20 g each of granulated mixture were withdrawn every 15 min from the fluidizer to quantify various parameters. The reference parameters thus determined were used to develop quantitative PLS models.

Moisture Content

The moisture content of the samples was determined by using the loss on drying (LOD) method.

This involved desiccating the samples in a oven at 115°C for 15 min. The resulting weight loss was taken to be the original moisture content of each sample.

Particle Size Distribution (PSD)

Each aliquot was passed through 9 sieves spanning the mesh range 1000–38 µm. The proportion of sample retained on each sieve was used to obtain the particle size distribution (PSD) curve for each aliquot. Three particle size ranges were defined as the specifications of this formulation (<125 µm, 125–1000 µm, and >1000 µm).

Bulk Density

The bulk density of the product was determined by weighing a known volume of granulate held uncompacted in a measuring cylinder.

Acquisition of NIR Data

Spectra were acquired with a QualitySpec ASD NIR spectrometer governed via the software Indico Pro v. BL 5.2 and equipped with three different detectors to span the whole wavelength region from 350 to 2500 nm, namely: a 512-photodiode array for the range 350–1000 nm and two InGaAs for the ranges 1000–1800 and 1800–2500 nm. All measurements were done on the NIR range (1000–2500 nm). The spectrometer is coupled to a optical probe for the spectral acquiring in reflectance mode. The experimental set-up included a 50 W halogen lamp and a holder that were placed around the glass window of the fluidizer (Fig. 1). All spectra were on-line acquired from the sample in continuous movement into the fluidizer. No spectrum was acquired from the sample in static mode outside the fluidizer. Preliminary tests involved variable positioning of the lamps and probe provided similar spectra. The position that provided better signal to noise ratio and a better reproducibility of the measurements was placing the optical probe closer to the window (less than 1 cm and 90° from the surface) and the lamp (5 cm, 45° from the surface). Each spectrum was obtained every 1 min and corresponded to the average of 100 scans.

Processing of NIR Data

Spectra were exported to the software The Unscrambler v. 9.2 (Camo Process AS, Oslo,

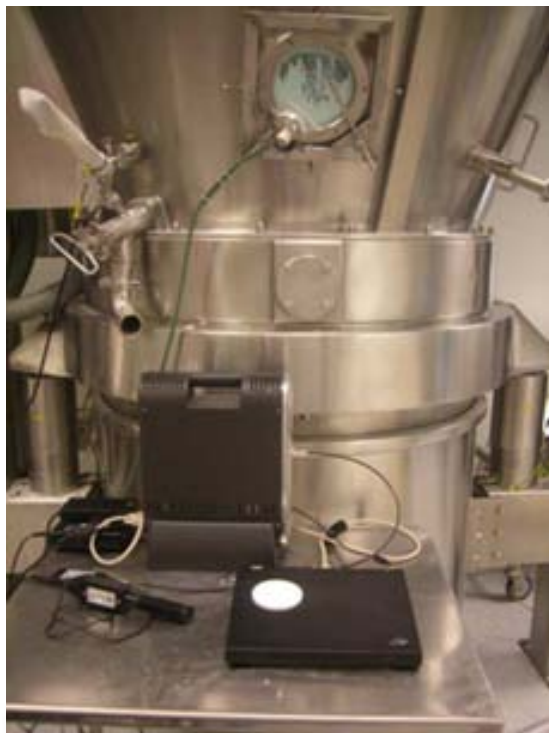


Figure 1. Fluidizer and NIR spectrometer.

Norway). The spectra obtained every 1 min intervals exhibited high variability as a result of their being recorded with the sample in continuous motion by effect of the circulating air stream. Moving average window of 7 spectra was used to decrease the spectral noise. The spectral pretreatments used to construct the calibration models included the standard normal variate (SNV), and the first and second derivatives. Derivative spectra were obtained by using the Savitzky–Golay algorithm using 11 points of moving smoothing window and a second-order polynomial.

Principal component analysis (PCA) was used to evaluate the spectral data sets of each individual process to obtain qualitative methods. Partial least squares (PLS) calibration models for the determination of moisture content (one calibration model for the percentage of water), particle size distribution (three calibration models, one for each percentage of retained sample through each sieve range: <125 μm , 125–1000 μm , and >1000 μm) and bulk density (one calibration model) were calculated using the spectral set corresponding to the intervals where aliquots were withdrawn and reference values were obtained. Calibration and prediction sample sets were selected encompassing the maximum spec-

tral variability (through scores plot of a PCA) and covering the entire value range for each determination. PLS calibration models were constructed by cross-validation using the leave-one-out method. The optimum number of factors was taken to be that leading to a minimum in the PRESS versus PLS component graph, the former being defined as:

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

where \hat{y}_i and y_i correspond to the reference and predicted value of each parameter for each sample (n), respectively. The goodness of calibration and prediction was assessed in terms of the Root Mean Square Error:

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

which was termed RMSEC for calibration and RMSEP for prediction. In all cases, the calibration models were calculated using the on-line NIR spectra acquired during the granulation process and the off-line reference values (particle size, moisture, or density). The obtained calibration models were applied to noninvasively monitor the granulation process quantifying the particle size, moisture and density.

RESULTS AND DISCUSSION

Spectra from five different processes monitored throughout via absorbance measurements over the wavelength range 1000–2500 nm have been used in the current study. Figure 2 illustrates the variation of the absorption bands during a single process. As can be seen, a significant increasing on the baseline is observed between before and after adding the binding solution. This spectral variation can be related with the increasing of particle size during granulation. Also, the bands in the regions of 1450 and 1950 nm increased markedly upon addition of the binder during the granulation step. Then, the second major spectral variation is observed at 1940 nm (O–H combination band), where the intensity clearly decreased at the end of the drying step. The major signal in the short-wavelength region (1000–1300 nm) was due largely to light absorption by the glass window. Two studies (qualitative and quantitative) have been carried with the whole spectral

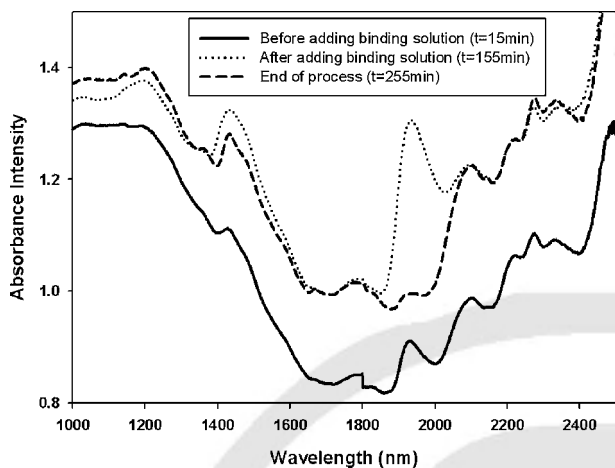


Figure 2. NIR spectra corresponding to three different steps of the granulation process.

set. Initially, the spectral changes during the processes with no additional information have been examined by PCA. This approach was intended to provide qualitative information of each step of the granulation process. Secondly, PLS calibration models were calculated for quantifying the target parameters of the formulation (viz. moisture content, particle size distribution, and bulk density).

Qualitative Study

A PCA of the body of absorbance spectra obtained for one of the processes revealed that the first two PCs in combination accounted for 99% of the total variance. Figure 3 is a PC1 versus PC2 scores plot for the process concerned as a function of the specific step. The scores in the left of the figure correspond to the initial, warming step; as can be seen, the cluster grouped together and exhibited no change. Upon addition of the binder solution 15 min later, both PC1 and PC2 scores changed for 55 min, after which only PC2 continued to vary. In the final step (drying) PC2 scores changed markedly and PC₁ only slightly. Figure 3 therefore clearly exposes the three steps of the granulation process.

The spectral and score changes in all studied processes were very similar to those of Figures 2 and 3. Therefore, a scores plot allows one to check whether a process is developing under normal operating conditions (NOC) and granulation is being properly done. PC1 accounted for 86% of the total variance; therefore, its associated parameter

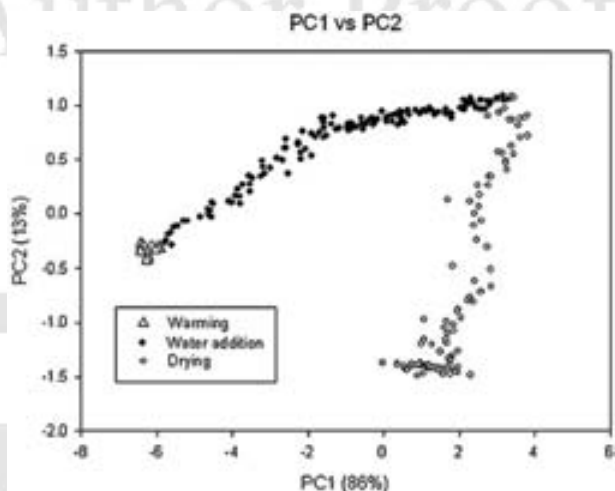


Figure 3. Scores plot obtained from a principal component analysis (PCA) of the complete spectral data set for a single process in absorbance (1000–2500 nm).

was the most influential on the process outcome. Figure 4a shows the variation of the PC1 scores with the time. The scores plot exhibits three distinct regions coinciding with as many steps: sample agitation and heating, addition of the binder solution and the granulate drying. As can be seen from their temporal variation, the scores remained constant during the first step (sample warming, which was the source of variation). This result can be explained as the spectra did not change or the spectral variation was insignificant. Therefore, the sample could be considered as homogeneous by the end of this step, but no drug determination was done.

The slight increase in the scores during the second step (addition of the binder solution) can be ascribed to the granulation operation; thus, the associated changes in particle size caused a spectral shift which was especially marked at the beginning and more gradual then. Also, the slight decrease in the scores (related with the variation of the spectra during time) during the third step is consistent with the decrease in particle size resulting from collisions between particles, which reduced their average size. The loading (related with the spectral profile change of each principal component) for PC1 (Fig. 4a and b) varied very similarly to the average spectrum for a single process; thus, the correlation coefficient between the two was greater than 0.94, which confirms that the first factor was associated to the change in particle size because an increasing in the particle size is revealed as a larger baseline displacement of the spectrum.

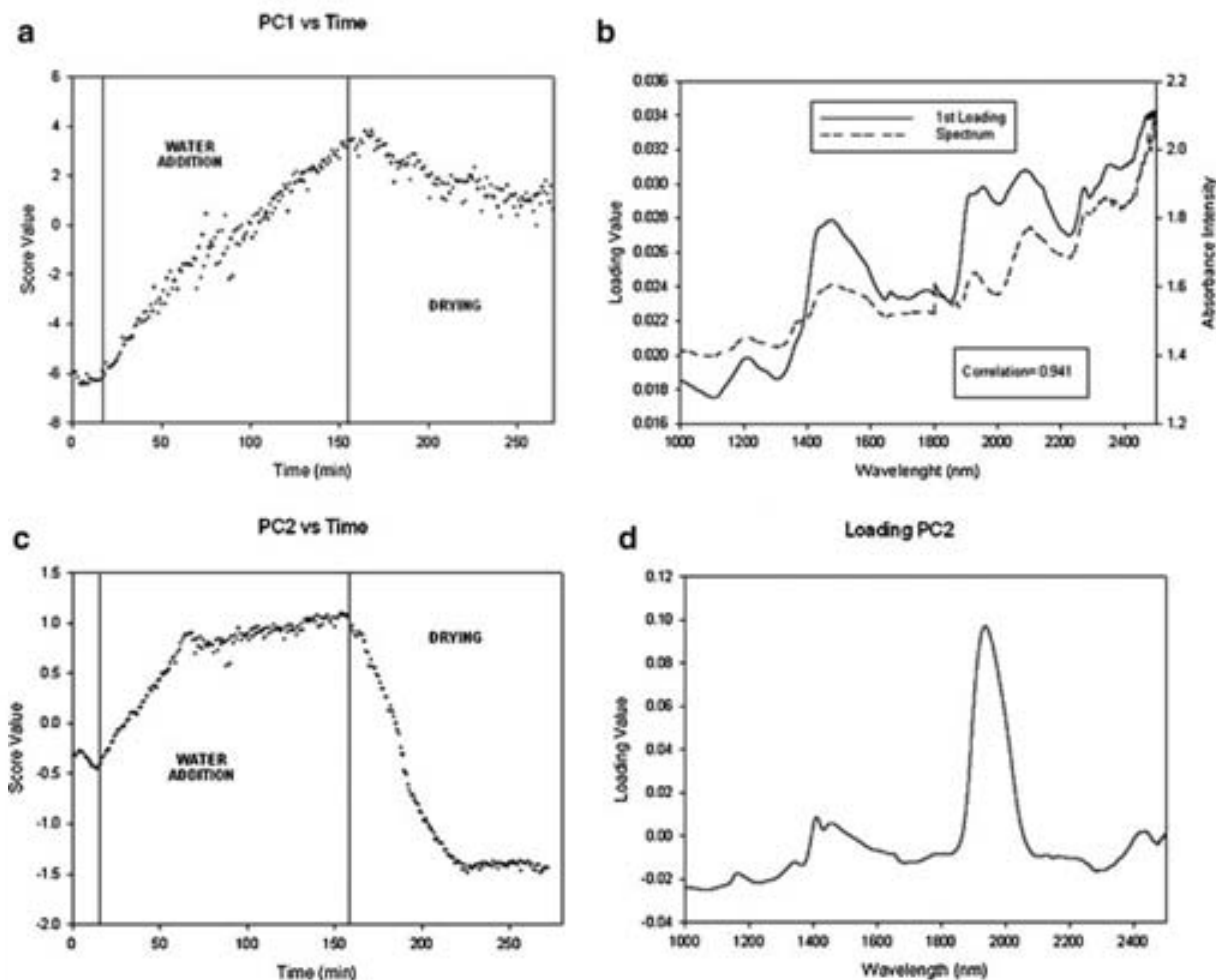


Figure 4. Scores and loadings plot from the PCA showed in (a) Scores PC1 versus time. b: Loading PC1 and mean spectrum of one process. c: Scores PC2 versus time, and (d) Loading PC2.

The variation of PC2 also gives interesting conclusions: the PC2 scores increased by effect of the binder addition in the second step and then decreased by effect of drying of the granules in the third (Fig. 4c and d). The weak decrease of scores values after 70 min (Fig. 4c) suggests the water content is reduced because the evaporation promoted by the hot air stream. Also, based on the decreased score, the amount of residual moisture in the end product was even smaller than that present at the start of the process. The variation of the loading for PC2 was also very similar to that of the water spectrum, indicating that this is the second most significant change in the spectra. Scores from PCA are related with spectral variation along the time, meanwhile the loadings can be compared with the spectral profile of a certain compound. In this case, the profile of

the second loading (Fig. 4d) is very similar to the spectrum of water. Therefore, the second factor must also be associated to water added during the granulation operation.

PC3 and PC4 were found to account for nearly 1% of the total explained variance; however, the temporal variation of their scores and loadings provided no additional information of use. Therefore, the variation of the PC1 and PC2 scores provides an effective knowledge of the process evolution and can be a useful tool for process analysis and control.

Quantitative Study

Approximately 20 samples every 15 min from each process were withdrawn. As the amount of binding solution increased (15–140 min), the

particle size of these aliquots was larger. Also, during the last step of the granulation (drying 140–240 min) the particle size decreased.

Five calibration models have been developed to predict the three target parameters (moisture content, particle size [three models] and bulk density) by using the reference values obtained from two of the five studied processes.

Particle Size

For production-oriented reasons, the particle size range 125–1000 μm was assumed to be suitable for monitoring with a view to facilitating subsequent compaction of the samples. However, also calibration models for the determination of particles smaller than 125 μm and larger than 1000 μm were formulated with the aim to be able to notice the amount of particles with undesirable size that generates segregation and clogging, respectively.

Figure 5 shows the variation in time of the results provided by the reference method for moisture content and fraction of sample from a granulation process compressed between 125 and 1000 μm sieves. This plot is similar to those of the variation of scores in Figure 4a and c (particle size with PC1 scores and moisture content with PC2 scores).

Particle size exhibited no appreciable change during the first step of granulation, which involved merely agitating the sample. However, the addition of the binder solution causes the formation of large granules and particle size to increase until the end of this second step. Finally,

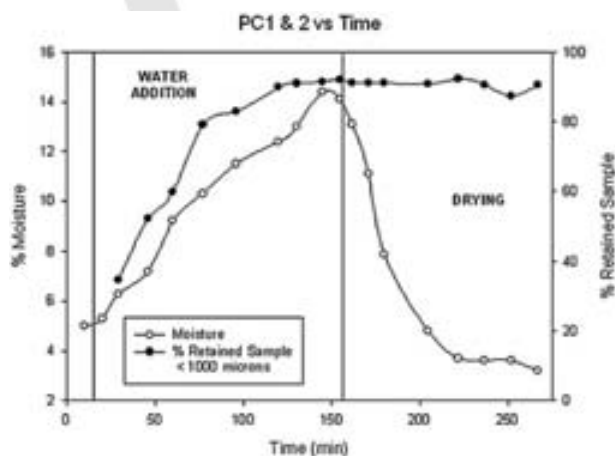


Figure 5. Evolution of moisture (%LOD) and percentage of retained sample lower than 1000 μm for one process.

the movement of the larger particles during the drying third step caused collisions between granules that reduced particle size and increased the fraction of sample retained on small-mesh sieves.

The variation of particle size during the wet granulation process was studied by using PLS models. Each withdrawn sample was passed through nine different sieves ranging in mesh size from 38 to 1000 μm . Samples were thus classified into three sets according to particle size: larger than 1000 μm , larger than 125 μm but smaller than 1000 μm , and smaller than 125 μm . Only 0.1% of sample passed through the tightest mesh (38 μm). The choice of sizes ranged was dictated by the need to obtain an end-product containing a high proportion of particles in their central region and a low proportion at the ends. The spectra for two of the processes were used to construct the calibration model and those for the others to validate it. In this way, one PLS1 calibration model for each of the three particle size ranges considered has been calculated (Tab. 1). In all cases, the calibration models were obtained in Absorbance mode which allows the better modeling between spectral variation and Y-values. Certainly, the effect of a change in particle size can be observed as a spectral baseline variation, so it is logic that better calibration models are obtained without applying spectral pretreatment. The number of PLS factors needed to construct each calibration models were four, five and seven (>1000 μm , 710–125 μm , and <125 μm , respectively). The higher number of PLS factors used on the <125 μm model can be related with the lower spectral variation observed when the particle size decreased. For the three calibration models, the prediction errors were between 1.4% and 7.8%.

Figure 6a–c illustrates the predictions for one of the processes, which spectra were not used to calculate the three PLS calibration models. In all cases, the predictive ability is demonstrated by the closeness between the reference values and the determinations of the process. No appreciable change in any particle size range was observed after 180 min; this suggests that granulation had completed by then. This was also the case with the scores for PC1, which can therefore be used to identify the end of the process.

Moisture

As revealed by the qualitative study (Fig. 4c and d), PC2 was associated to the moisture content of

Table 1. Figures of Merit of the Three Calibration Models for Determination of Particle Size Distribution

	>1000 μm	710–125 μm	<125 μm
Calibration set			
Spectral mode		Absorbance	
Spectral range (nm)	1000–1400	1000–1215	
	500–1875	1630–1900	1000–2500
	2100–2500	2150–2425	
PLS factors	5	4	7
Samples	21	21	20
% retained sample	1.04–10.73	34.88–96.21	2.17–21.22
% explained variance	97.9	97.5	97.6
NIR vs. REF	$0.979x + 0.111$	$0.975x + 1.98$	$0.963x + 0.55$
RMSEC (%)	0.60	2.79	1.45
Prediction set			
Samples	22	23	23
% retained sample	1.23–4.72	37.48–93.84	2.82–17.46
RMSEP (%)	1.40	7.78	4.56

the samples being granulated. This led us to develop a PLS1 model for quantifying moisture during the process. SNV spectral pretreatment was applied to the spectral matrix with the aim to reduce the effect of particle size. Similarly to the previous study, spectra from two processes were used to develop one calibration model and the predictive ability was tested with the other three processes. Table 2 shows the figures of merit of the ensuing PLS model. Although the first PLS factor accounted for 95% of the variance, a total of three factors were needed to ensure close fitting and a good predictive ability (prediction error of 0.93% and 99% explained variance).

Figure 6d shows the predicted variation of the moisture content as calculated by applying the model to one of the processes not used for calibration and the respective reference values for comparison. The moisture content of the granulate reached 3% after about 220 min and then leveled off.

Bulk Density

Similarly to the previous determinations, the reference values for bulk density were obtained from samples withdrawn during the studied processes. One PLS calibration model was has been calculated using spectra from two processes and the predictive ability was evaluated analyzing the spectra of the other three processes. The figures of merit of the resulting model are shown in Table 2. In this case, the calibration model obtained in Absorbance mode allowed the deter-

mination of the bulk density. The authors understand the calibration range is relatively narrow (0.325–0.406 g/mL) reason why a calibration model can be obtained with three PLS factors. However, the predictive ability of this calibration model is acceptable, obtaining a prediction error of 0.03 g/mL.

Figure 7 illustrates the predicted results for a process not included in the calibration model and the corresponding reference values. The bulk density of the product followed the same trends-occasionally of opposite sign-as its particle size and moisture content. The results of the determinations of withdrawn samples exhibited the same trend as those predicted by the model. Thus, the bulk density leveled off after about 220 min, beyond which no substantial changes in moisture content or particle size were observed. Since bulk density is directly related to moisture and particle size, changes during the process can be detected by monitoring these parameters.

CONCLUSIONS

The proposed noninvasive methodology allows on-line monitoring of a granulation process in a fluid bed granulator. Spectral data has been acquired using a noninvasive near infrared (NIR) spectroscopy optical probe attached to the window of the fluidizer. Qualitative method has been obtained using principal component analysis (PCA) applied to the spectral data. This method allowed to identify and monitor each step of the granulation

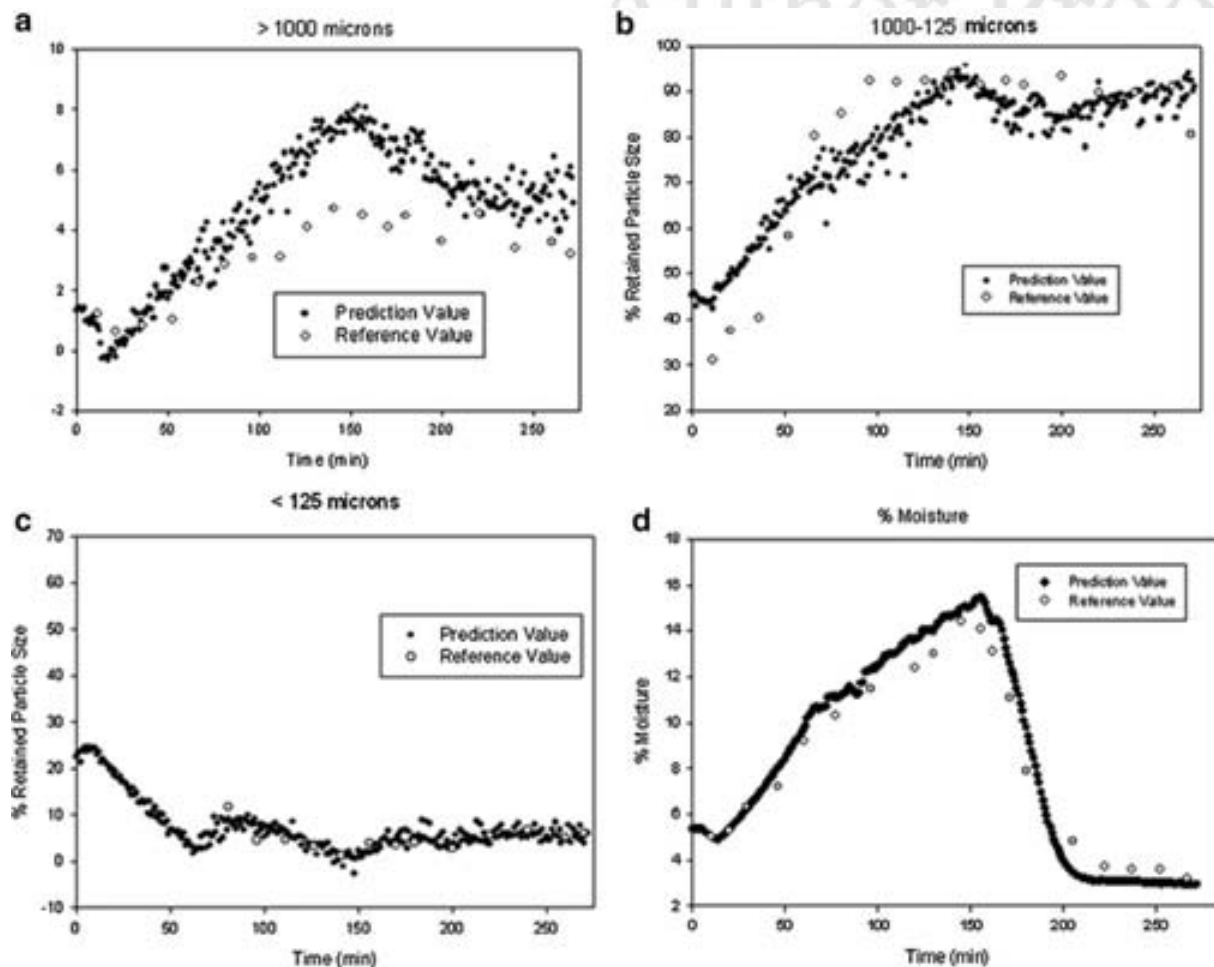


Figure 6. Prediction of particle size distribution and moisture content for one [process^{Q2}](#).

Table 2. Figures of Merit of the Calibration Models for Determination of Moisture Content and Bulk Density

	Moisture Content	Bulk Density
Calibration Set		
Spectral mode	Standard normal variate	Absorbance
Spectral range (nm)	1800–2000	1485–2375
PLS factors	3	3
Samples	31	22
Range		
Moisture (%) density (g/mL)	1.8–16.2	0.325–0.406
% Explained variance	98.9	97.6
NIR vs. REF	$0.989x + 0.093$	$0.976x + 0.008$
RMSEC	0.54%	0.01 g/mL
Prediction set		
Samples	31	37
Range		
Moisture (%) density (g/mL)	3.2–14.4	0.329–0.405
RMSEP	0.93%	0.03 g/mL

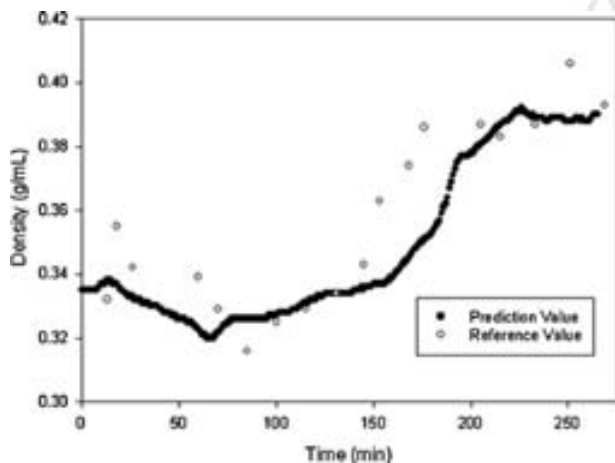


Figure 7. Prediction of bulk density and reference values for one process.

process by the interpretation of scores and loadings of the PCA. Both physical and chemical information can be obtained simply visualizing the spectra after principal component analysis. In addition, quantitative methods have been developed by the use of partial least squares (PLS) calibration technique. These calibration models allowed the determination of the relevant quality parameters of the formulation: moisture content, particle size distribution and bulk density. Both qualitative and quantitative methods were validated by using external reference data. In all cases, a correct predictive ability of the methods was confirmed.

ACKNOWLEDGMENTS

The authors thank the MCyT of Spain for funding this research within the framework of Project CTQ2006-12923, and Laboratorios Menarini, S.A. (Badalona, Spain) for funding this study.

REFERENCES

- Royce A, Mecadon M, Holinej J, Karnachi A, Valazza S, Wei W. 2005. Process control and scale-up of high shear wet granulation. *Am Pharm Rev* 2:8–16.
- Reich G. 2005. Near-infrared spectroscopy and imaging: Basic Principles and pharmaceutical applications. *Adv Drug Del Rev* 57:1109–1143.
- U.S., Department of Health and Human Services. Food and Drug Administration (2004) Guidance for industry: PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance. <http://www.fda.gov/cder/guidance/6419fnl.pdf>.
- ICH Guidelines Q8, Q9, Q10. <http://www.ich.org/cache/compo/276-254-1.html>.
- Rantanen J, Jorgensen A, Räsänen E, Luukkonen P, Airaksinen S, Raiman J, Hänninen K, Antikainen O, Yliruusi J. 2001. Process analysis of fluidized bed granulation. *AAPS PharmSci* 2:Q3.
- Otsuka M, Mouri Y, Matsuda Y. 2003. Chemometric evaluation of pharmaceutical properties of antipyrine granules by near-infrared spectroscopy. *AAPS PharmSci* 3: article 47.
- Sekulic S, Ward HW, Brannegan DR, Stanley ED, Evans CL, Sciavolino ST, Hailey PA, Aldridge PK. 1996. On-line monitoring of powder blend homogeneity by near-infrared spectroscopy. *Anal Chem* 68:509–513.
- Rantanen J, Wikström H, Turner R, Taylor LS. 2005. Use of in-line near-infrared spectroscopy in combination with chemometrics for improved understanding of pharmaceutical processes. *Anal Chem* 77:556–563.
- Nieuwmeyer FJS, Damen M, Gerich A, Rusmini F, Maarschalk KV, Vromans H. 2007. Granule characterization during fluid bed drying by development of a near infrared method to determine water content and median granule size. *Pharm Res* 24:1854–1861.
- Rambali B, Baert L, Massart DL. 2003. Scaling up of the fluidized bed granulation process. *Int J Pharm* 252:197–206.
- Herting MG, Kleinebudde P. 2007. Roll compaction/dry granulation: Effect of raw material particle size on granule and tablet properties. *Int J Pharm* 338: 110–118.
- Rambali B, Baert L, Massart DL. 2001. Using experimental design to optimize the process parameters in fluidized bed granulation on a semi-full scale. *Int J Pharm* 220:149–160.
- Vojnovic D, Chicco D, El Zenary H. 1996. Doehlert experimental design applied to optimization and quality control of a granulation process in a high shear mixer. *Int J Pharm* 145:203–213.
- Frake P, Greenhalgh D, Grierson SM, Hempenstall JM, Rudd DR. 1997. Process control and end-point determination of a fluid bed granulation by application of near infra-red spectroscopy. *Int J Pharm* 151:75–80.
- Green RL, Thurau G, Pixley NC, Mateos A, Reed RA, Higgins JP. 2005. In-line monitoring of moisture content in fluid bed dryers using near-IR spectroscopy with consideration of sampling effects on method accuracy. *Anal Chem* 77:4515–4522.
- Macias K, Carvajal MT. 2008. An assessment of techniques for determining particle size during high-shear wet granulation. *Tablets Capsules* 6: 32–40.

**6.5 STUDY OF PHARMACEUTICAL SAMPLES BY
NIR CHEMICAL IMAGE AND MULTIVARIATE
ANALYSIS**

J.M.Amigo, J.Cruz, M.Bautista, S.Maspoch, J.Coello,
M.Blanco.

Trends in Analytical Chemistry. Volume 27, Issue 8,
September 2008, Pages 696-713

Study of pharmaceutical samples by NIR chemical-image and multivariate analysis

José Manuel Amigo, Jordi Cruz, Manel Bautista, Santiago Maspocho,
Jordi Coello, Marcelo Blanco

Near-infrared spectroscopy chemical imaging (NIR-CI) is a powerful tool for providing a great deal of information on pharmaceutical samples, since the NIR spectrum can be measured for each pixel of the image over a wide range of wavelengths.

Joining NIR-CI with chemometric algorithms (e.g., Principal Component Analysis, PCA) and using correlation coefficients, cluster analysis, classical least-square regression (CLS) and multivariate curve resolution-alternating least squares (MCR-ALS) are of increasing interest, due to the great amount of information that can be extracted from one image. Despite this, investigation of their potential usefulness must be done to establish their benefits and potential limitations.

We explored the possibilities of different algorithms in the global study (qualitative and quantitative information) of homogeneity in pharmaceutical samples that may confirm different stages in a blending process. For this purpose, we studied four examples, involving four binary mixtures in different concentrations.

In this way, we studied the benefits and the drawbacks of PCA, cluster analysis (K-means and Fuzzy C-means clustering) and correlation coefficients for qualitative purposes and CLS and MCR-ALS for quantitative purposes.

We present new possibilities in cluster analysis and MCR-ALS in image analysis, and we introduce and test new BACRA software for mapping correlation-coefficient surfaces.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Augmented-MCR-ALS; BACRA; CLS; Correlation coefficient; Fuzzy C-means clustering; K-means clustering; MCR-ALS; NIR image; PCA; Surface homogeneity

1. Introduction

Chemical imaging (CI) or, more specifically, near-infrared CI (NIR-CI) promises to become one of the most powerful techniques for providing high-quality spectral information on surfaces. The robust, reliable combination of chemical (molecular spectroscopy) and physical (digital imaging) features generates increasing interest in many different applications [1–4].

NIR-CI is especially attractive in pharmacology [1,5–7], where it offers enough information for non-destructive studies of important parameters (e.g., homogeneity or polymorphism) for the whole surface of a pharmaceutical tablet simultaneously over a range of spatial resolution. NIR-CI has been applied in a number of different pharmaceutical fields (e.g., drug release in the dissolution of tablets [8–10], uniformity of contents [11], or the study of

granulation process with mapping techniques [12]).

NIR-CI is also a scanning technique that could be introduced into new process analytical technology (PAT) for the control of pharmaceutical processes, where obtaining information about quality is essential and robust methodologies for analyzing the datasets are always in demand [1,7,13–17].

The pharmaceutical processes for making tablets usually comprise several steps (e.g., blending, granulation, and compression). One important step is the blending stage of all the components to ensure that the correct formulation is achieved. Obtaining reliable information about the development of this stage of the process is essential in assuring the quality of the final product.

An excellent review recently covered general aspects of CI, specifically including

José Manuel Amigo*

Department of Food Science,
Faculty of Life Sciences,
University of Copenhagen,
Rolighedsvej 30, DK-1958
Frederiksberg-C, Denmark

**Jordi Cruz, Manel Bautista,
Santiago Maspocho,**

Jordi Coello, Marcelo Blanco
Unidad de Química Analítica,
Departament de Química,
Facultat de Ciències,
Universitat Autònoma de
Barcelona, E-08193 Bellaterra,
Barcelona, Spain

*Corresponding author.
E-mail: jmar@life.ku.dk

CI applied to pharmaceutical process monitoring [18]. This article, besides being an overview of the NIR-CI technique, reviews applications of several chemometrics methods not only in NIR-CI, but also in other fields of hyperspectral CI (Raman imaging).

In this context, several methodologies (e.g., principal component analysis (PCA), correlation coefficients, cluster analysis, partial least square regression-discriminant analysis (PLS-DA), linear discriminant analysis (LDA), classical least squares (CLS) and multivariate curve resolution-alternating least squares (MCR-ALS)) have previously been reported and have demonstrated their usefulness for exploratory analysis of the surface [19–25].

A priori, these methods can be classified in two well-established groups:

- those aimed at offering mainly qualitative information (exploratory analysis) (e.g., PCA, Fixed Size Image Window-Evolving Factor Analysis (FSIW-EFA), cluster classification, LDA and its variant (PLS-DA)), and correlation coefficients; and,
- those expected to give an estimation of the concentration of the analytes involved in each pixel (e.g., CLS and MCR-ALS).

All these methods have their own features, advantages and disadvantages that have to be taken into consideration when setting a real objective in pharmaceutical research. All of the methods mentioned have generally been applied where the main target has been to distinguish and/or group the sample into its main components. Nevertheless, a more interesting problem arises when, apart from the influence of the pure signal of each component, it is also important to take into consideration the possible non-homogeneous distribution of the components in the mixture. This problem is frequently encountered in the blending process for pharmaceutical tablets, where assurance of good mixing of the different analytes is essential in order to meet the regulatory standard for the tablet.

We have divided this work into two distinct parts:

- First, we describe the state of the art and the latest trends in multivariate methods in pharmaceutical research. We also include some basic theoretical background to show the main features and the drawbacks of the mathematical implications of each method.
- Second, we offer some perspectives on the practical possibilities and the limitations of the main multivariate techniques in the study of the evolution of blending processes by NIR-CI (e.g. PCA, K-means (KM) clustering, Fuzzy C-means (FCM) clustering, correlation coefficients, CLS and MCR-ALS).

To achieve this, we present new perspectives on the use of clustering and MCR-ALS in samples with lack of selectivity (homogeneous mixture), and we introduce and test new software for the study of correlation-coef-

ficient surfaces of mixtures, known as BACRA (Basic Analysis of Correlation between Analytes).

For this study, we selected four examples of binary mixtures of ibuprofen and starch, which could be used as model in any blending process. The selection of ibuprofen and starch was significant. A solid pharmaceutical preparation generally comprises two components: the so-called active principle ingredient (API) and the placebo (containing all the excipients). Ibuprofen is a well-known API in a number of pharmaceutical preparations, and starch can represent the placebo. Perfect knowledge of the distribution of components of samples helps to clarify different concepts and to state conclusions about the usefulness of the different methods tested.

2. Structure of hyperspectral data

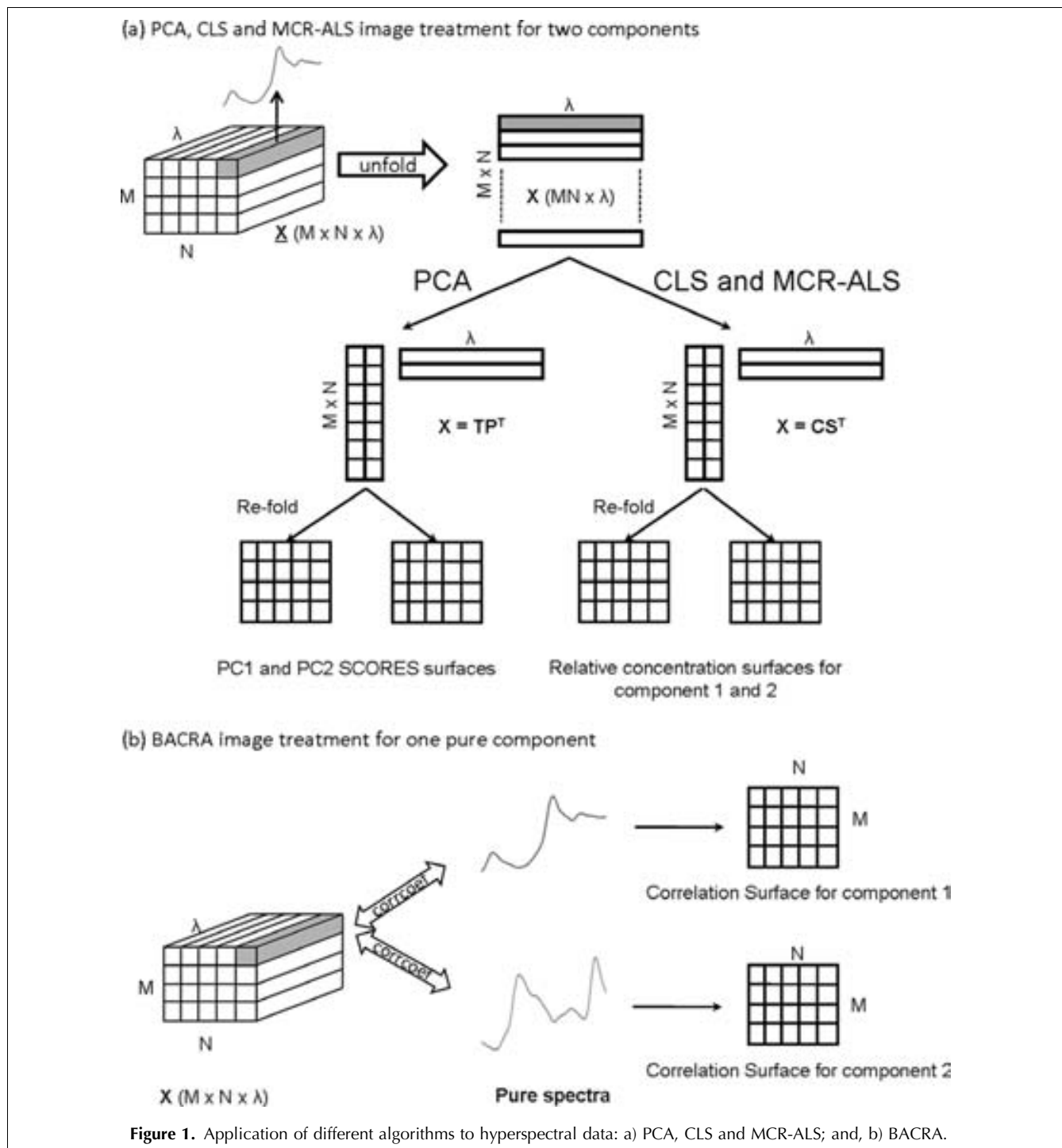
One approach to multivariate analysis in NIR-CI was the use of single and representative wavelength mapping and further multivariate analysis [26,27]. We have focused on the multivariate analysis of the images by using the whole wavelength range since the usefulness of using more than a group of specific wavelengths has been demonstrated [27].

As we have commented upon, one of the main features of measurements with any hyperspectral imaging methodology is the great amount of information collected for one sample (e.g., in NIR-CI, one NIR spectrum can be measured over a wide wavelength range in each pixel of the image). The result is a 3-D data set, known as a spectral hypercube (Fig. 1). For the structure of \mathbf{X} ($M \times N \times \lambda$), the M and N axes represent spatial information and the λ axis corresponds to the spectral pattern. This is a 3-D data cube [22], but it is also common to call this array a hyperspectral image [23].

Such data structures can comprise thousands of spectra, so it is evident that this raw data, normally very correlated with each other, needs powerful mathematical treatments to extract the desirable information.

In NIR-CI, it is important to assess whether the spectra collected were chemically representative of the sample measured [28], so a calibration stage is needed for the NIR-CI device in order to make effective use of the images. We do not intend to go deep into all the methods to calibrate NIR-CI devices. Nevertheless, some studies developed in this context [29–31] have shown the importance of the previous calibration of the NIR-CI device to obtain chemical information and to avoid artifacts (e.g., originating from light sources and shadows).

For a global analysis of the hyperspectral image, the well-known 3-D PARAFAC algorithm (parallel factor analysis) [32] could be used. Nevertheless, hyperspectral images are far from the trilinearity condition, since each pixel may have a different concentration profile. This



situation means that PARAFAC, or even PARAFAC2 [33,34], cannot analyze hyperspectral images.

3. Preprocessing the hyperspectral image

Before any mathematical preprocessing, a crucial step is unfolding the spectral hypercube. The unfolding step is essential to adapt the 3-D array for further preprocessing

or applying any 2-D algorithms (e.g., PCA, clustering, CLS and MCR-ALS). In essence, when $\underline{\mathbf{X}}$ ($M \times N \times \lambda$) is unfolded, the two-way matrix obtained, \mathbf{X} , will have the dimensions ($MN \times \lambda$) (Fig. 1a).

The preprocessing step is essential in order to avoid the influence of undesirable phenomena classically encountered in NIR measurements [35], and the benefits can be directly extrapolated from classical NIR analysis, where data preprocessing is performed to facilitate extraction of

chemical information in subsequent data-analysis procedures by attenuating undesirable signals arising from particle-size effects or morphological differences, such as surface roughness and detector artifacts.

Different methods (e.g., smoothing/derivatives (Savitzky-Golay or Norris algorithms), Standard Normal Variate (SNV) [36], Multiplicative Scatter Correction (MSC) [37] and normalizing to unit-vector length) are frequently used in image-data analysis. One question regarding the interpretability of results is that minimal preprocessing is preferred, since preprocessing may introduce data artifacts in several ways. For this reason, SNV and smoothing are preferable, since they preserve as much as possible the spectral shape, reducing undesirable noise in the spectral profile and pixel-to-pixel intensity differences.

SNV is a very common pre-treatment in NIR spectroscopy to eliminate variations in the baseline promoted by light scattering. SNV transforms each spectrum to a spectrum with a mean-intensity value of zero and unit standard deviation. An important consideration is that SNV operates on each individual spectrum independently.

Smoothing techniques allow elimination of the noise level of the spectra obtained. This is a common practice in NIR techniques that helps to minimize the variability of the data related to noise. Most commonly, smoothing uses Savitzky-Golay methodology, which employs a polynomial smoothing window around a central point to calculate smoothing value for that window. Two parameters need to be chosen for the calculation: the degree of the polynomial and the window width. Generally, polynomials of degree 2 or 3 perform similarly [26]. The choice of an optimal window size is important and some optimization is often necessary to choose the optimal window size [38,39].

4. Techniques for exploratory analysis

There are already options available for the exploration of pharmaceutical tablets (e.g., wavelets transformation [40], previous PCA or FSIW-EFA [23], which is a variant of the well-known EFA [41] for hyperspectral images). The basis of FSIW-EFA relies in developing PCA consecutively to specific areas of the image.

4.1. Principal Component Analysis (PCA)

PCA decomposes the unfolded \mathbf{X} matrix as follows (Fig. 1):

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (1)$$

where \mathbf{T} is the $(MN \times F)$ scores matrix, \mathbf{P} is the $(\lambda \times F)$ loading matrix and \mathbf{E} is the $(MN \times \lambda)$ residual matrix for F principal components (PCs) (Fig. 1a).

In essence, PCA is variable reduction technique that condenses all the spectral information into a few latent variables, the PCs. Each score vector can be re-folded to show the relative distribution map for each PC (Fig. 1a). In this sense, the loading vectors in combination with each score surface can be used to relate each PC with any variability source in the sample. This way of decomposing the hyperspectral image is commonly also known as the Multiway-PCA algorithm (MPCA) [42–44].

PCA has been previously used in hyperspectral imaging mainly for the study of texture [40], homogeneity [5,9,20,22,45] and locating contaminants [46] on the surface of tablets, as well as for assessment of quality parameters in the process of manufacturing tablets [9,47].

PCA has two main drawbacks that may make it unattractive for hyperspectral imaging. The first one is related to calculation of the “chemical rank” of the matrix. The chemical rank of a hyperspectral image can be defined as the variability sources that make the signal significantly different from the noise. In an ideal situation, the “chemical rank” of the hyperspectral image should match the number of components in the sample. Nevertheless, in a PCA model, the PCs do not necessarily correspond to the chemical rank of the matrix, since each PC is obtained by maximizing the amount of variance. In this sense, it is worth noting the widespread discussion about the applicability of PCA in hyperspectral imaging. While some authors claim (arguably) that they are able to assign the first 18 components to each pure analyte in one sample [48], other authors demonstrate that PCA is not a good technique for features extraction [24]. This misunderstanding is explained by the fact that, effectively, PCA is mainly a technique for searching variability sources, taking into consideration that the variability explained for one PC can be associated to a pure component, to a mixture or, what is more important, to a physical effect (e.g., light scattering or incorrect position of the sample).

Another problem with PCA models is the visual interpretation of the scores and loadings. Each PC is the vector product of its score and its loading, as shown in Eq. 1. This decomposition can also be represented as follow:

$$x_{mn\lambda} = t_{mn1}p_{\lambda 1}^T + t_{mn2}p_{\lambda 2}^T + t_{mn3}p_{\lambda 3}^T + \dots + t_{mnf}p_{\lambda f}^T + e_{mn\lambda} \quad (2)$$

where $x_{mn\lambda}$ represents each $mn\lambda$ th of the \mathbf{X} unfolded image, t_{mnf} and $p_{\lambda f}$ correspond to the score and loading profiles for each f th PC. Nevertheless, any vector product may be rotated and, instead of $t_{mnf}p_{\lambda f}^T$, we can have the product $(-t_{mnf})(-p_{\lambda f}^T)$ (Equation 3):

$$x_{mn\lambda} = (-t_{mn1})(-p_{\lambda 1}^T) + t_{mn2}p_{\lambda 2}^T + t_{mn3}p_{\lambda 3}^T + \dots + t_{mnf}p_{\lambda f}^T + e_{mn\lambda} \quad (3)$$

It is evident that the mathematical results of PCA are the same in Equations 2 and 3. Nevertheless, this mathematical property may cause problems in elucidating the sources of variability that each PC is explaining when a direct visual interpretation is needed.

The combined use of the correlation coefficients between the loading profiles and the pure spectra (if they are available) may help us guess whether or not any PC is rotated.

4.2. Cluster analysis

Cluster analysis conforms a number of classification methods that perform the analysis of the image by segmenting the data into specific chemical composition based on the information of the pixels. The main feature of clustering methods is that they classify objects (in NIR hyperspectral images, the objects are the pixels) into groups without involving a previous learning set [22].

Before cluster analysis, the number of variables should be reduced to a sufficient minimum while the maximal amount of information is retained (feature extraction). This improves computational efficiency and reduces the complexity of the cluster model. Good options are to use wavelet transform [49] or the scores obtained in a previous PCA as input [22].

Among the different hierarchical and non-hierarchical cluster algorithms used in image analysis [50,51], we will focus on two non-hierarchical algorithms for cluster analysis (KM and FCM).

4.2.1. K-means algorithm. The KM algorithm assigns each pixel x_{mn} of the image to the k th cluster, whose center is nearest, by minimizing the sum of the squared distances of each pixel to its corresponding center (Equation 4):

$$J = \sum_{k=1}^K \sum_{mn=1}^{MN} \|x_{mn}^{(k)} - m_k\|^2 \quad (4)$$

where $x_{mn}^{(k)}$ is the mn th pixel and m_k is the center of cluster k . The algorithm steps are [52]:

- (1) Choose the number of clusters k
- (2) Generate k clusters and determine the cluster centers
- (3) Assign each pixel to the nearest cluster center
- (4) Calculate J and recalculate new cluster centers
- (5) Repeat 3 and 4 till convergence criterion.

The main advantage of this algorithm is the simplicity that allows it to run with larger datasets. Nevertheless, the main drawbacks of the KM algorithm are: there is a risk of converging to a local minimum in the iterations; and, it forces to each pixel to belong exclusively to one cluster.

4.2.2. Fuzzy C-means algorithm. In FCM clustering, each pixel is assigned a fractional degree of membership to all the clusters simultaneously, rather than it belonging completely to one cluster (as in KM clustering). In this sense, FCM allows pixels in the edge to belong to one cluster to a lesser degree than pixels that are in the middle of the cluster. In FCM clustering, membership function coefficient u_{mnk}^g is calculated for each x_{mn} pixel in such a way that each coefficient is compressed between 0 and 1, and the sum of all the coefficients is defined to be 1 (Equations 5 and 6):

$$u_{mnk}^g = \left[\sum_{i=1}^C \left(\frac{\|x_{mn}^{(k)} - m_k\|^2}{\|x_{mn}^{(k)} - m_i\|^2} \right)^{1/(g-1)} \right]^{-1} \quad (5)$$

$$\forall x_{mn} \rightarrow \sum_{k=1}^K u_{mnk}^g(x_{mnk}) = 1(6)$$

where g is the so-called “fuzzifier” constant, which determines the fuzziness of the clustering result [22]. A good value of g is 2, which indicates that the coefficients are linearly normalized to make this sum 1.

Now, the J function is calculated as follows:

$$J = \sum_{k=1}^K \sum_{mn=1}^{MN} u_{mnk}^g \|x_{mn}^{(k)} - m_k\|^2 \quad (7)$$

and cluster center m_k is calculated as the weighted mean:

$$m_k = \frac{bf \sum_{mn=1}^{MN} u_{mnk}^g x_{mn}}{\sum_{mn=1}^{MN} u_{mnk}^g} \quad (8)$$

The algorithm steps for FCM are very similar to those for KM [53]. The main differences are the fuzzed assignation of each pixel and that FCM also minimizes intra-cluster variance. Nevertheless, the same problems of local minima and finding the appropriate number of clusters are encountered in FCM.

4.2.3. Number of clusters

Usually in CI, the word “cluster” is understood as a group of pixels in the global sample that are defined by the same analyte. In pharmaceutical-blending processes, the number of components is usually perfectly known beforehand. This seems to solve the big problem of clustering methods: finding the appropriate number of clusters. Nevertheless, the problem not only remains, but it increases in importance in the sense that each sample being blended comprises the pure components and a number of linear combinations between pure components (concentration gradients). So the original definition of cluster has to be “re-formulated”. In pharmaceutical-blending processes, one cluster can be considered as an entity of the sample comprising a number of pixels whose spectral profiles differ significantly from the main components. In this sense, the pure components define the classes of the sample and there are as many clusters as many different linear combina-

tions between the components whose spectral profiles differed significantly from the spectral profiles of the classes.

Several empirical indexes have already been proposed in the literature to calculate the number of clusters [54]. In this work, we focus on Silhouette index (for KM) and Partition Entropy (PE) index (for FCM).

4.2.3.1. Silhouette index. The Silhouette index [54] is calculated for each x_{mn} pixel and offers a measure about the similarity between points in the same cluster compared to points in other clusters (Eq. 9):

$$(S_{mn})_k = \frac{b_{mn} - a_{mn}}{\max(a_{mn}, b_{mn})} \quad (9)$$

where a_{mn} is the average distance between each m th pixel and all the pixels included in the same cluster and b_{mn} denotes the minimum average distance between each m th pixel and the pixels included in other clusters. S_{mn} values close to 1 denote that this pixel has been correctly associated to the clusters; when the S_{mn} value is close to -1 , there is a "misclassification" of this pixel. A more visual way to compare Silhouette indexes is to calculate the average Silhouette index for each k number of clusters:

$$\text{mean}(S_{mn})_k = \frac{\sum_{mn=1}^{MN} S_{mn}}{MN} \quad (10)$$

4.2.3.2. Partition Entropy index. The PE index [55] for FCM can be defined as:

$$PE = \frac{-1}{MN} \sum_{k=1}^K \sum_{mn=1}^{MN} u_{mnk}^g \log(u_{mnk}^g) \quad (11)$$

where the natural logarithm is used for weighted coefficient u_{mnk}^g and MN denotes the total amount of pixels of the image. The PE value ranges from 0 to $\log(K)$. Values close to 0 indicate a good estimation of the number of clusters, whereas PE values close to $\log K$ indicate that the number of clusters does not reflect the real structure of the image.

Most times, the interpretation of Silhouette and PE indexes is not evident and they do not offer a clear estimation of the number of clusters in the sample. In this sense, if the raw data are used to develop KM or FCM cluster analysis, mean centre m_k is expected to have the shape of the mean spectral profile of all the pixels grouped in that cluster. The comparison between mean centre profile and pure analyte profile may help to elucidate how well the clustering analysis has been developed and that the algorithm has not converged to a local minimum. Nevertheless, if the mean centers are highly similar each other, it can suggest that an excessive number of clusters have been used.

Re-sampling is a good way to test reliability and robustness of the cluster model. It comprises the following steps:

- (1) Randomly choose a representative amount of pixels of the image
- (2) Calculate the PE index or the Silhouette index for one number of clusters
- (3) Repeat steps 1 and 2 a certain number of times and calculate the mean value and the standard deviation of the PE index or the Silhouette index.

These three steps have to be repeated as many times as the number of clusters are studied. If the standard deviation increases with the number of clusters, this may indicate an inconsistent number of clusters and fewer clusters may be chosen.

4.3. Similarity using correlation coefficients

Correlation coefficient is a well-known parameter for measuring the similarity between spectra [5]. In this case, correlation coefficients are calculated between pure spectra of the analytes and each spectra obtained from corresponding single pixel in the sample (Fig. 1b), obtaining a correlations map that shows visually the distribution of each analyte in the sample. The correlation coefficient is calculated as follow:

$$\text{corrcoef} = \frac{\sum_{\lambda} x_{mn\lambda} y_{\lambda}}{\sqrt{\sum_{\lambda} x_{mn\lambda}^2 \sum_{\lambda} y_{\lambda}^2}} \quad (12)$$

where, for each wavelength, $x_{mn\lambda}$ is the spectrum measured in the m th pixel and y_{λ} is the pure spectrum profile for one analyte. The highest similarity is achieved between two identical spectra, when the coefficient is 1 [5]. Correlation coefficients are basically qualitative parameters that may help obtain a real qualitative overview about the distribution of each analyte in the image.

5. Techniques for estimating analyte concentration in each pixel

Several techniques have been applied to obtain an estimation of the concentration of analytes in each pixel in NIR-CL. Two groups can be differentiated: those that require a previous stage of performing a calibration model; and, those that do not strictly need a previous calibration set of images.

Among the first group, apart from PC Regression (PCR) [56], the most useful multivariate methodology is PLS Regression (PLSR) [5,6,40,46,57,58] and its variant PLS2 [59,60]. Despite their demonstrated applicability, the previous stage of calibration may make PLSR unattractive when fast analysis of samples is needed.

For this reason, we want to focus on techniques that estimate the concentration in each pixel without needing a previous stage of calibration: CLS and MCR-ALS.

Between both groups is PLS-DA, which is a classification technique that needs a previous stage of

calibration. Nevertheless, with PLS-DA, it is possible to perform the calibration model using only pure spectra, so, in this situation, PLS-DA is analogous to CLS.

5.1. Classical Least Squares

Classical Least Squares regression (CLS) is a very well-known method of quantification that assumes that each measurement is the weighted sum of linearly independent signals (e.g., in NIR-CI systems, CLS assumes that, in the absence of interactions within the components of the sample, the spectrum of each pixel is the sum of the all pure-component spectra weighted by their concentration [5,9,59]).

In essence, the unfolded \mathbf{X} matrix is decomposed as follow:

$$\mathbf{x} = \mathbf{cS} \quad (13)$$

where \mathbf{x} is the measured response vector, \mathbf{c} is the vector containing the concentration and \mathbf{S} is the pure-component spectrum. The real interest in CLS lies in that, when all the pure spectra that constitute the mixture are perfectly known, \mathbf{c} can be easily calculated as follows:

$$\mathbf{c} = \mathbf{xS}^+ \quad (14)$$

where S^+ denotes the pseudo-inverse matrix of S :

$$S^+ = S^T(SS^T)^{-1} \quad (15)$$

Despite the ease of calculating this concentration matrix, CLS has two main drawbacks:

- the signal can be affected by other non-systematic variations that influence the spectral pattern; and,
- all pure spectra are not always available.

For these reasons, another more feasible method is required.

5.2. Multivariate Curve Resolution-Alternating Least Squares

The so-called Self-Modeling Curve Resolution (SMCR) techniques present an interesting alternative to CLS when the pure spectra are not known. Some authors have already used SIMPLe-to-use Interactive Self-Modeling Mixture Analysis (SIMPLISMA) to study homogeneity in pharmaceutical samples [22]. Probably, the most widespread SMCR method is MCR-ALS [61], which has already been shown to be a very useful tool for local and global analysis of hyperspectral images in combination with a previous technique for exploratory analysis (e.g., PCA or FSIW-EFA) [22,23].

MCR-ALS decomposes matrix \mathbf{X} ($MN \times \lambda$) into the product of two matrices, \mathbf{C} ($MN \times F$), containing the concentration profiles, and \mathbf{S}^T ($F \times \lambda$), containing the spectral profiles for each F component (Equation 16):

$$\mathbf{X} = \mathbf{CS}^T + \mathbf{E} \quad (17)$$

where \mathbf{E} ($MN \times \lambda$) accounts for the experimental error. MCR-ALS works by iteratively optimizing the matrices \mathbf{C}

and \mathbf{S}^T with the help of constraints based on chemical knowledge or on mathematical features of the data.

Basically, MCR can be compared to CLS, but with several differences. With MCR, there is no need for the pure spectra, and the calculation of \mathbf{C} and \mathbf{S} matrices is performed iteratively. Several algorithms have been proposed to solve this product, being ALS the most widespread.

ALS needs initial estimates to start the iterations. Working with good initial estimates can help the algorithm to converge to a good solution, avoiding local minima. For example, a good starting point is to use pure spectra, if they are available. A stopping criterion has to be imposed. The most common is the percentage of Lack Of Fit (%LOF) that indicates the difference between input data \mathbf{X} and the data reproduced from the \mathbf{CS}^T product for each iteration [62].

The main drawback of MCR-ALS is that the resolution of the product \mathbf{CS}^T is not so evident, since this product has not a unique solution. Nevertheless, this problem can be minimized by including several soft-modeling constraints. The most useful in Pharmaceutical imaging are:

- non-negativity in concentration and/or spectral profiles, which imposes that concentration profiles of the components are always supposed to be positive; and,
- closure, at which each pixel can be supposed to accomplish a constant mass balance of 1, representing 100% of global concentration.

Before starting MCR-ALS decomposition, the number of components has to be determined. Several methods have been described to solve this question. A good option for hyperspectral imaging is the FSIW-EFA method developed by de Juan et al. [23]. This PCA-based method for screening the rank of hyperspectral images across small intervals of the image shows the variation in the number of compounds needed to describe each of the pixel windows in the hyperspectral image.

5.3. Augmented MCR-ALS for homogeneous samples

One drawback of MCR-ALS dealing with homogeneous images is the lack of selectivity, which can promote problems of ambiguity. Variability in homogeneous samples is expected to be due to only one component, the mixture of several analytes. This variability means that the sample does not have any area selective to any pure component. This lack of selectivity can be associated to a rank-deficiency problem.

This problem may be minimized by matrix-augmentation analysis. Matrix augmentation analyzes several samples simultaneously, profiting from the relationships of one to the others [63]. The main feature of matrix augmentation is that the homogeneous sample has to be augmented with an image without lack-of-selectivity problems (full-rank image).

The performance of image augmentation in a column-wise manner is the same as in 2-D matrices, but includes a previous step of unfolding the 3-D image (Fig. 2):

$$\begin{pmatrix} X \\ D \end{pmatrix} = \begin{pmatrix} C_X \\ C_D \end{pmatrix} (S^T) + \begin{pmatrix} E_X \\ E_D \end{pmatrix} \quad (18)$$

where X ($MN \times \lambda$) is the full-rank image and D ($AB \times \lambda$) is the rank-deficient image. C_X and C_D are the concentration surfaces simultaneously obtained for each image and S^T is the transpose of the global spectral profiles. E_X and E_D are the residual matrices.

By modeling the rank-deficient image in this column-wise augmented mode, the property of full-rank of X image is improved and ambiguities associated with the lack of selectivity of the D matrix are minimized.

6. Experimental and data treatment

6.1. Reagents and instruments

Ibuprofen and starch were of analytical grade (Kern Pharma, Terrasa, Spain). Four homogeneous mixtures of

ibuprofen and starch were prepared, with 75%, 65%, 50% and 25% ibuprofen, respectively. Three vessels were split in three parts and one of them was filled by the binary samples of 75%, 50% and 25%, respectively. The other two parts were filled with pure components (Fig. 3). A fourth vessel was totally filled with the mixture of 65% of ibuprofen, simulating the final blended sample.

6.2. Experimental

The chemical hyperspectral images of the four samples were collected by diffuse reflectance mode using an NIR-CI spectrometer [Think Spectrally Roda-25 (Think Spectrally, Valencia, Spain)] with a mercury-cadmium telluride (MCT) detector, which provided a spatial resolution of 320×256 pixels. The size of each pixel was $40 \times 40 \mu\text{m}$, giving an imaging area of $10 \times 12 \text{ mm}$. Spectra were measured with 10 scans, in a spectral

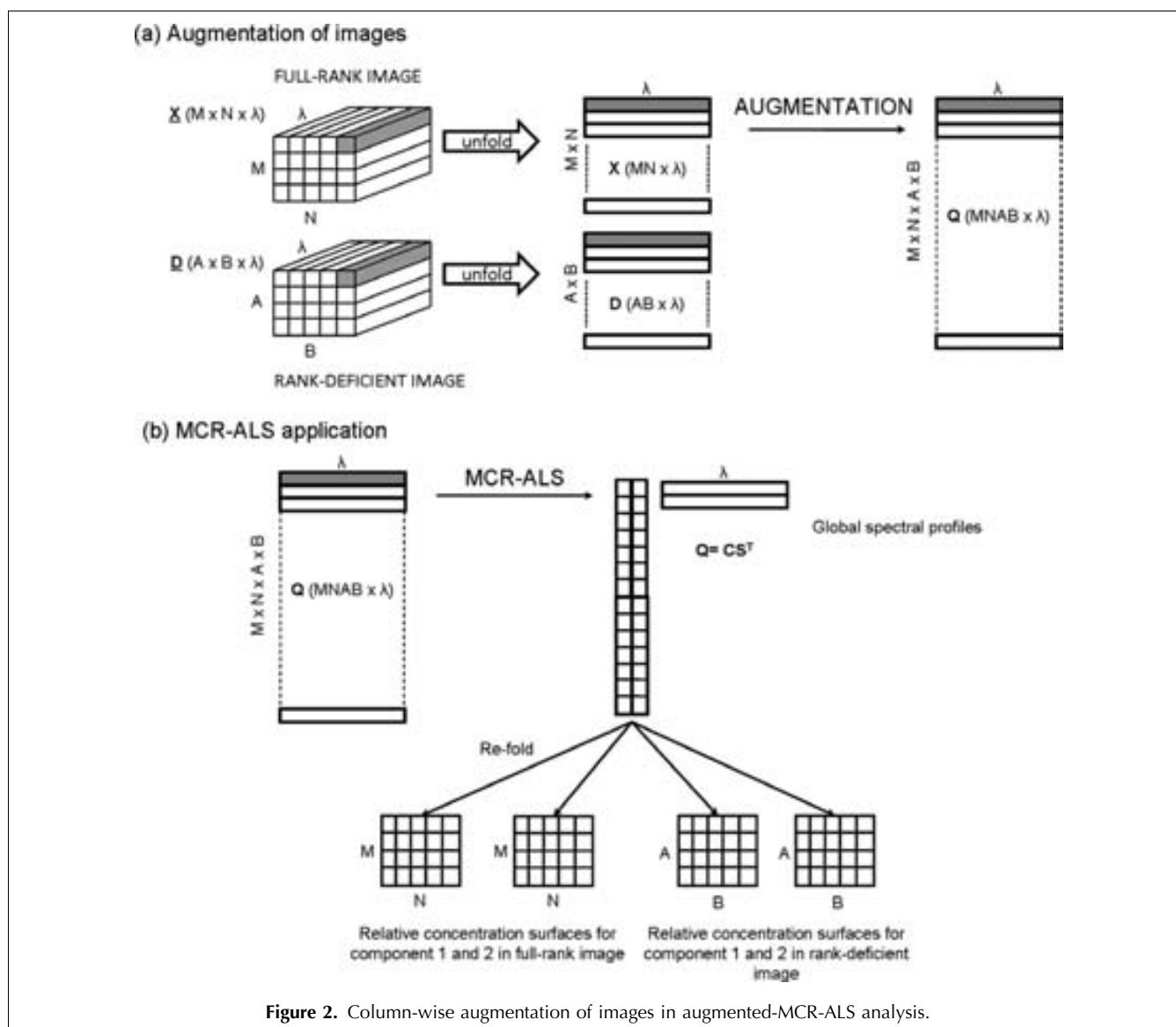


Figure 2. Column-wise augmentation of images in augmented-MCR-ALS analysis.

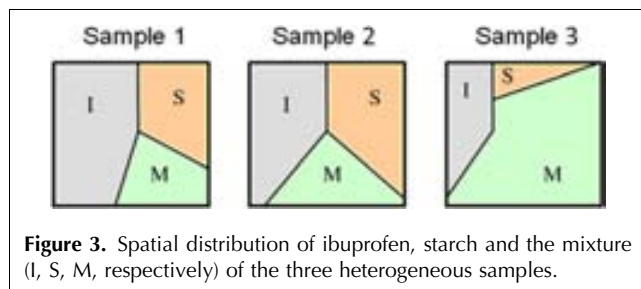


Figure 3. Spatial distribution of ibuprofen, starch and the mixture (I, S, M, respectively) of the three heterogeneous samples.

6.3. Data treatment

Savitzky-Golay smoothing of the spectral pattern was applied with a window size of 11 and a second polynomial order. SNV was also applied to the four images.

Correlation-coefficient surface mapping, PCA, CLS, Savitzky-Golay smoothing and SNV algorithms were implemented in BACRA, working under in MatLab v. 7.5 (MathWorks, Massachusetts, USA). Apart from pre-processing, PCA and CLS results, BACRA allowed visualization of the spatial distribution of several compounds in a mixture using the correlation coefficients calculated between the NIR spectrum obtained for each pixel and the spectra of the pure compounds. A correlation map for each analyte in the sample and histograms of correlation coefficients for each analyte were therefore used to study the distribution of the analytes in the measured surface. The software and detailed, easy-to-follow instructions are freely available from the authors or via the Internet [64].

For MCR-ALS and augmented-MCR-ALS applications, the software provided by Tauler et al. was used [62,65]. The value of %LOF of 0.5% was used in this work, because of the high variability of the NIR images.

Clustering algorithms implemented in Statistical Toolbox and Fuzzy Logic Toolbox (MathWorks) were also used, as well as Silhouette and PE indexes. The re-sampling method was only applied to FCM clustering and programmed in MatLab code.

7. Results and discussion

7.1. PCA analysis

PCA results for the four samples are shown in Fig. 4. The explained variance for each PC was also plotted (Fig. 4). For comparison, the correlation coefficients between the obtained loading profiles and pure spectra of ibuprofen and starch were also calculated (Table 1).

From Fig. 4, it can be observed that the first two PCs explain more than 99.8% of the total variance. We could try to associate each of these PCs to each of the pure components of the samples. Nevertheless, this assump-

tion may lead to wrong conclusions about the PCA results.

For example, it is easy to see that results for Sample 1 denote that the loading of PC1 is highly correlated to the spectrum for pure ibuprofen (Table 1). Nevertheless, the surface of the score for that PC (Fig. 4a) does not show a clear difference between the area with pure ibuprofen and the area with mixture (containing 75% ibuprofen). This is because the high correlation with pure ibuprofen that this area presents (Table 1).

The opposite situation with respect to Sample 1 is encountered in Sample 3. The loading of the PC1 has a higher correlation with starch and mixture area than with ibuprofen (Table 1). Now the PC1 score surface (Fig. 4c) explains the area with pure starch and mixture (containing 75% of starch).

For Sample 2, the results are still more confusing. The correlation values of the first loading with ibuprofen and starch spectra are quite similar (Table 1). Also the correlation values of the mixture area are quite similar for both analytes. In this sample, the mixture area contains the same amount of both analytes.

Fig. 4d shows PCA results for Sample 4. The first PC explains 99.96% of the total variance, whereas the second PC only explains 0.02%. This indicates a homogeneous mixture with a variability that may be explained with only one PC. The plot of the first loading profile seems to be a linear combination of ibuprofen and starch spectra (results not shown). The correlation coefficients between the first loading profile and pure analytes corroborate this; they are very similar and present high values, as seen in Table 1.

As has been demonstrated, it is very dubious to assign each PC to one pure component, even in these easy-to-follow samples, and only when the contribution to the variability explained by the PC differs for both analytes there is a difference between them in the image of the scores.

All these results illustrate what was pointed before: a PC do not necessarily correspond to one specific chemical component.

An important fact is observed in the score surfaces of Sample 4 (Fig. 4d). At first sight, two different areas can be distinguished in the score surface for PC2. Nevertheless, the loading profiles do not clarify the meaning of this fact. Further investigation with other methods has to be done to obtain an explanation.

7.2. Cluster analysis

7.2.1. K-means results for heterogeneous samples. The mean Silhouette index calculated for each cluster in KM analysis is shown in Fig. 5a, indicating that two clusters for Samples 1 and 3 seem to result in a feasible solution (Figs. 6a and c). Curiously, these are the samples in which the mixture-area concentration is closer to those for areas containing pure

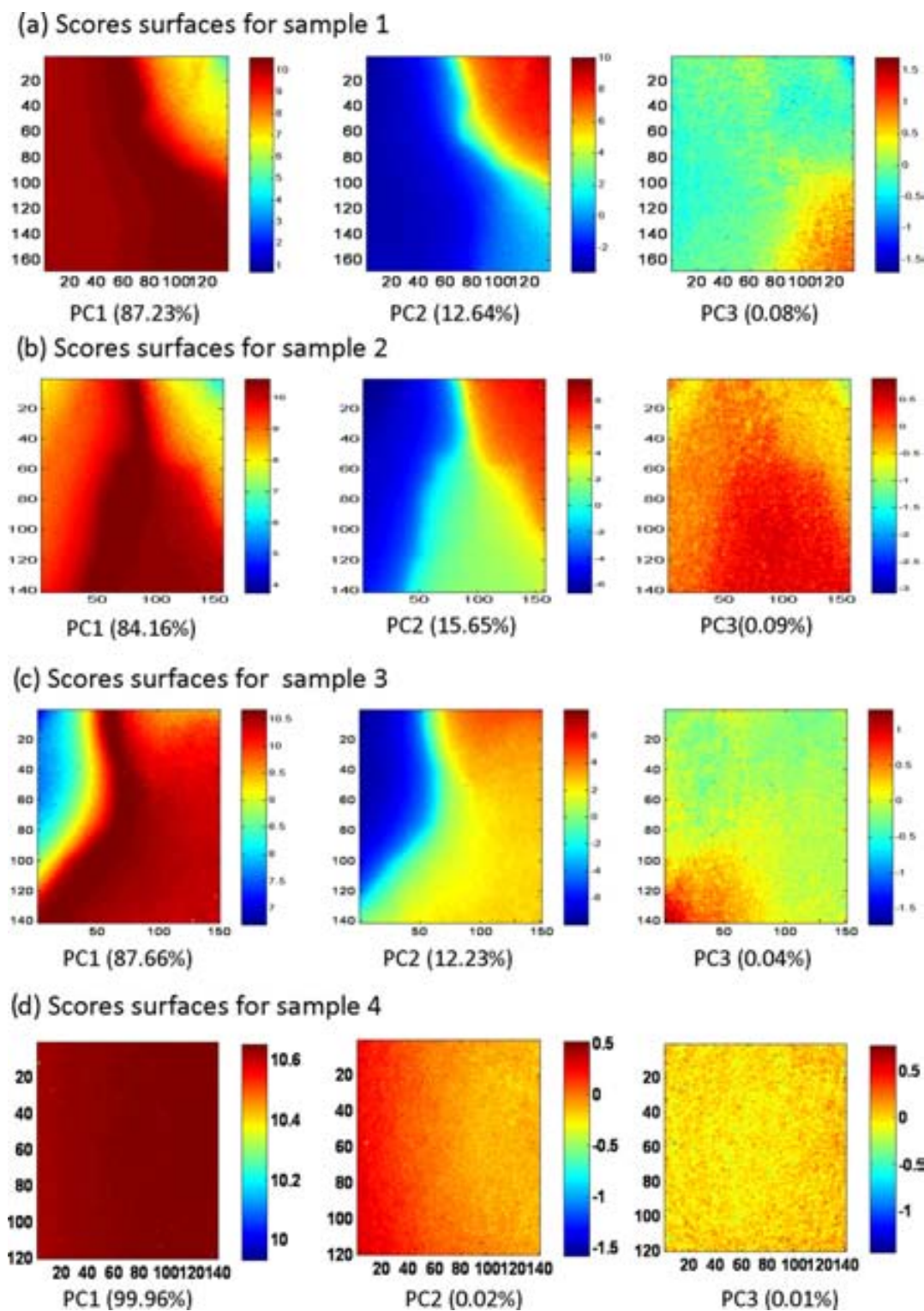


Figure 4. PCA analysis. False-color pictures of the first three scores surfaces for a) Sample 1; b) Sample 2; c) Sample 3; and, d) Sample 4. Cold (blue) and hot (red) colors refer to low and high intensities, respectively.

analyte. In Sample 1, the mixture area was associated with the area if pure ibuprofen. By contrast, the mixture area of Sample 3 was associated with the area of pure starch.

The shapes of the centers for these two clusters are almost identical to those of their corresponding pure analytes (correlation coefficient above 0.989, results not shown). This result indicates that there are not enough

		Ibuprofen	Starch
Sample 1	Loading 1	0.9343	0.5919
	Loading 2	-0.3506	0.8032
	Loading 3	0.0151	0.0152
	Raw data ^a	0.9254	0.5806
Sample 2	Loading 1	0.7993	0.7914
	Loading 2	-0.5956	0.6076
	Loading 3	0.0155	0.0196
	Raw data ^a	0.7067	0.8567
Sample 3	Loading 1	0.6398	0.9105
	Loading 2	-0.7634	0.408
	Loading 3	0.0678	0.027
	Raw data ^a	0.466	0.9697
Sample 4	Loading 1	0.9096	0.8294
	Loading 2	0.2471	0.2748
	Loading 3	0.0018	0.0974
	Raw data ^b	0.9095	0.8314

^aCorrelation calculated between the pure spectrum and the mean spectrum of the mixture area.
^bCorrelation calculated between the pure spectrum and the mean spectrum of the sample.

spectral differences between mixture areas and the corresponding area of pure analyte to separate them into different clusters.

The Silhouette index of the second sample suggests that three clusters explain the surface much better (Fig. 5a). The distinction between the areas containing pure analytes and the mixture area was possible because of the concentration of the mixture area (50% of each component), making its spectra significantly different from the spectra of pure analytes.

Despite this explanation of the results for the different areas of the samples, KM analysis is not feasible enough to analyze these kinds of samples where there are mixture areas containing a linear combination of pure analytes.

7.2.2. FCM results. As stated above, FCM methods are more feasible, since the degree of belonging to the different clusters is calculated for each pixel.

A clear number of clusters has not been obtained by calculating the PE index using a re-sampling methodology (the index for each number of clusters was calculated 100 times, using a 25% of the pixels each time randomly). The PE index for each sample (Fig. 5b) increases in a linear way with the number of clusters; this means that each cluster is explaining an equivalent part of the information. Moreover, the standard deviation remains constant in each sample, assessing the robustness of the mean values of PE obtained.

Considering that each sample is composed of both pure compounds and an undefined number of linear combinations between them, and knowing that, at least, there are two classes clearly defined, FCM analysis was performed considering only two clusters.

FCM analysis of two clusters is shown in Figs. 7a, b and c. The shape of the cluster centers obtained corresponds perfectly with the pure spectra of ibuprofen and starch (correlation coefficients above 0.980). Now, a surface has been obtained for each cluster in the samples. Looking at Fig. 7, these surfaces explain the theoretical distribution of the analytes.

Taking into consideration previous results, the FCM analysis of the homogeneous sample considering two clusters was performed (Fig. 7d). The surfaces seem to explain that there are two well-distributed clusters.

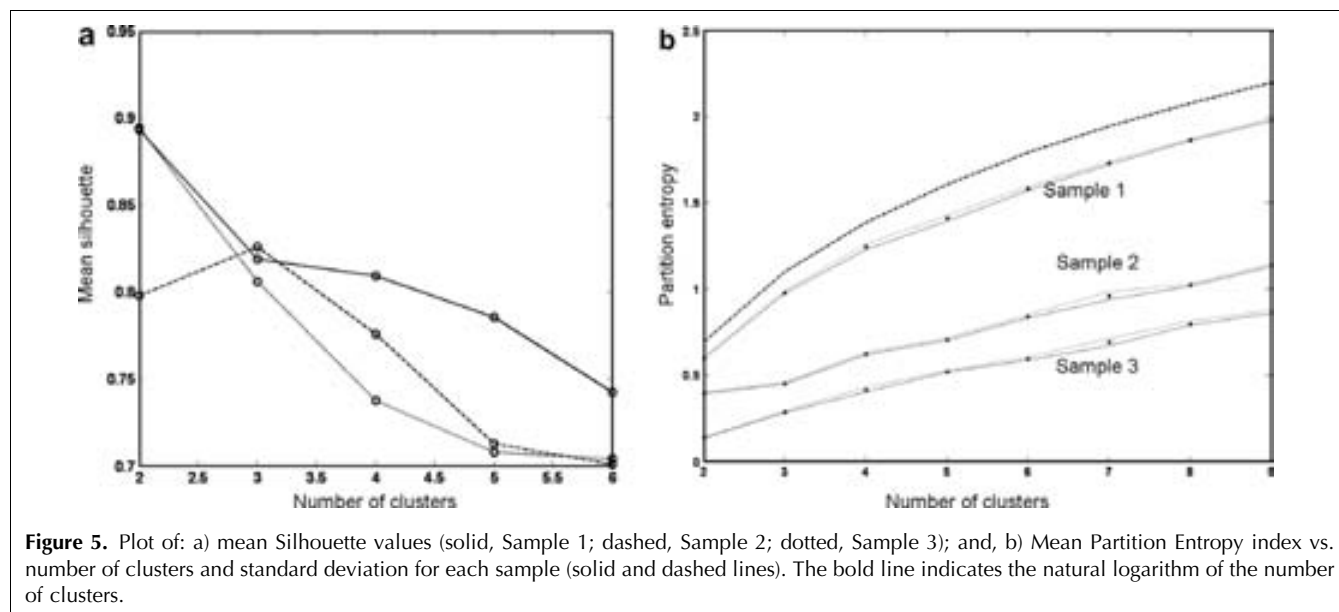


Figure 5. Plot of: a) mean Silhouette values (solid, Sample 1; dashed, Sample 2; dotted, Sample 3); and, b) Mean Partition Entropy index vs. number of clusters and standard deviation for each sample (solid and dashed lines). The bold line indicates the natural logarithm of the number of clusters.

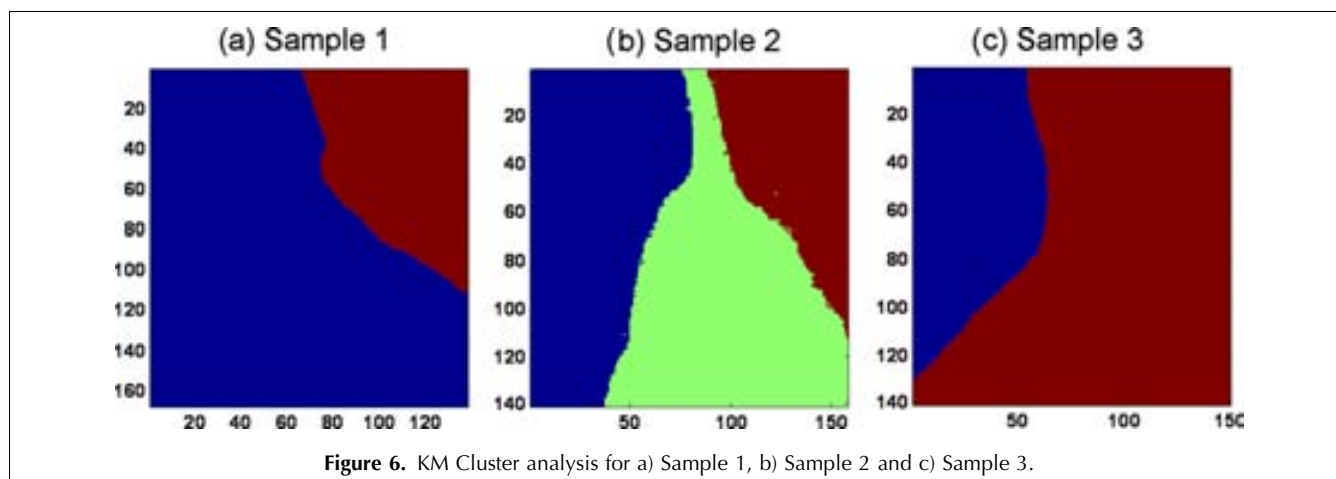


Figure 6. KM Cluster analysis for a) Sample 1, b) Sample 2 and c) Sample 3.

Nevertheless, the centers obtained for each cluster (Fig. 7) have very similar shapes (correlation = 0.999). This indicates that both clusters are explaining almost the same information. The very small differences between both centers may indicate that, assuming homogeneity, there exists an area reflecting variation in the light or the surface was not totally horizontal. This result helps to clarify the PCA results where the existence of this area was denoted (Fig. 4d), but not totally explained.

7.3. Correlation-coefficient maps–BACRA results

The results obtained for the three heterogeneous samples are shown in Figs. 8a, b and c, respectively, where the histograms of the correlation coefficients clearly confirm that the samples were not homogeneous. For the three samples, the correlation-coefficient maps for ibuprofen distinguish the zones, from highest to lowest correlation, according to the spatial distribution of the analyte in the sample.

Looking at the correlation-coefficient maps in the mixture zones, the intensity of the color changes from Sample 1 (reddish color) to Sample 3 (light green-yellow color), indicating that the correlation between the spectra obtained in these areas with the pure spectrum of ibuprofen is decreasing sample to sample in line with the decreasing of ibuprofen contents.

The correlation coefficient between pure ibuprofen and pure starch is 0.298. In the histogram of ibuprofen for Sample 1 (Fig. 8a), there is a near-constant number of low correlation-coefficients value from this minimum to about 0.80 that we assume to belong to the area with no or little ibuprofen; the number of pixels then increases when the correlation-coefficient values approach 1, which corresponds to the mixture area and the area of pure ibuprofen.

When the amount of ibuprofen decreases in the mixture area, the distribution of the correlation coefficients moves to lower values. As can be expected, the opposite situation occurs for starch, suggesting that there is a

relationship between the correlation coefficients and the concentration of the analytes in the mixtures.

Looking at the histograms of correlation coefficients for Sample 2 (Fig. 8b), it can be seen that the group of coefficients related to the mixture area for ibuprofen appears around 0.70, whereas the group for starch appears around 0.90 (green circle in Fig. 8b). This indicates that the correlation between the spectra of the sample and the pure spectrum of starch is higher than the correlation with pure spectrum of ibuprofen. This situation can also be seen in the false-color images of the correlation-coefficient maps. This result can be attributed to the higher absorbance of starch with respect to ibuprofen, so there is a higher contribution to the mixture signal from part of the starch.

Results for the homogeneous sample are shown in Fig. 8d. The histograms of the correlation coefficients calculated between the image of the sample and the pure components fit a Gaussian distribution for ibuprofen and starch, offering a reliable qualitative parameter of the homogeneity of the sample. The correlation maps show a homogeneous distribution in the centre of the sample, but being slightly different in the left of the image. This appreciable match with the results of FCM clusters indicates a slight deviation of the spectral profiles in this area.

7.4. CLS results

The concentration surfaces for ibuprofen and starch follow the expected trend for Samples 1, 2 and 4, but not for Sample 3 (results not shown). The areas with pure ibuprofen and pure starch are perfectly quantified but not the concentration obtained in Sample 3 nor in the mixture areas for all the samples, with average values far away from the real ones (Table 2). Nevertheless, these values have only an indicative meaning, since only a small area of the mixture was considered.

The mean concentrations obtained for ibuprofen and starch in the homogeneous mixtures are 65% and 48%,

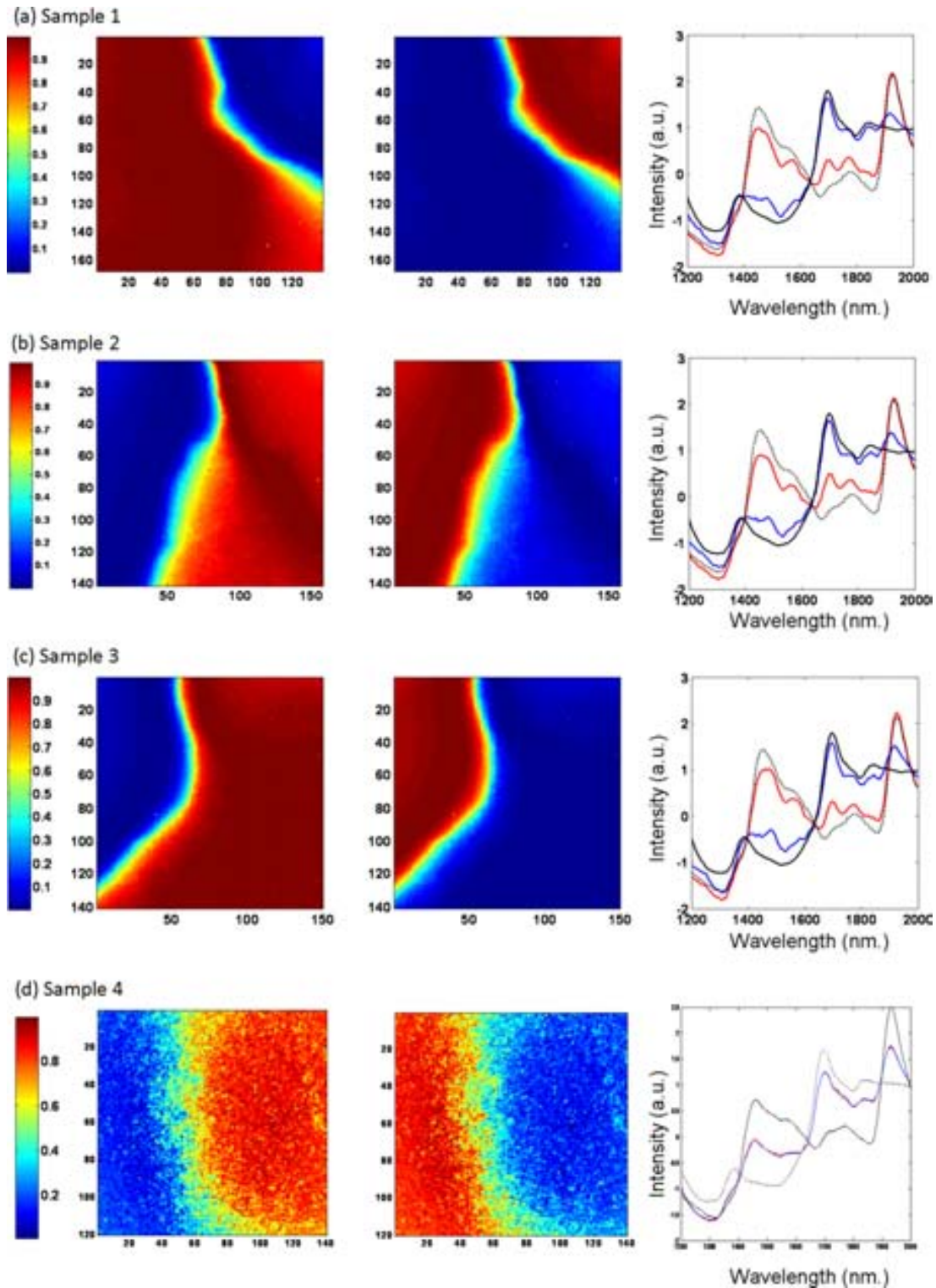
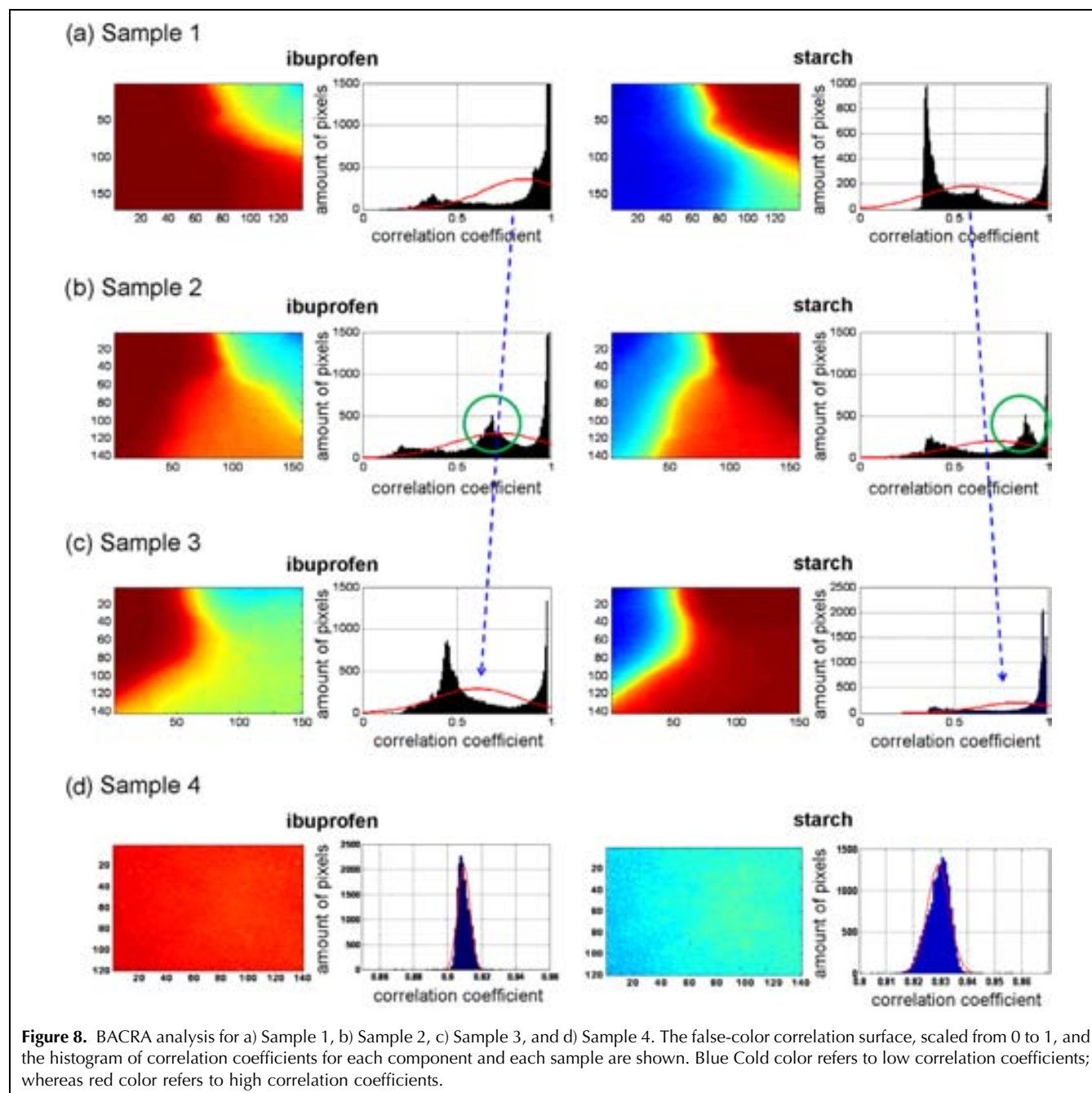


Figure 7. FCM Cluster analysis of the four samples and the corresponding centers of each cluster (solid black, spectrum of pure ibuprofen; dashed black, spectrum of pure starch).

respectively (Table 2). The concentration calculated for ibuprofen perfectly matches the theoretical one. Nevertheless, the concentration for starch is far from that

expected. This bad result could be because the spectral profiles obtained are affected by other sources of variability that do not affect the pure spectra.

**Table 2.** Results obtained for CLS, MCR-ALS and augmented-MCR-ALS

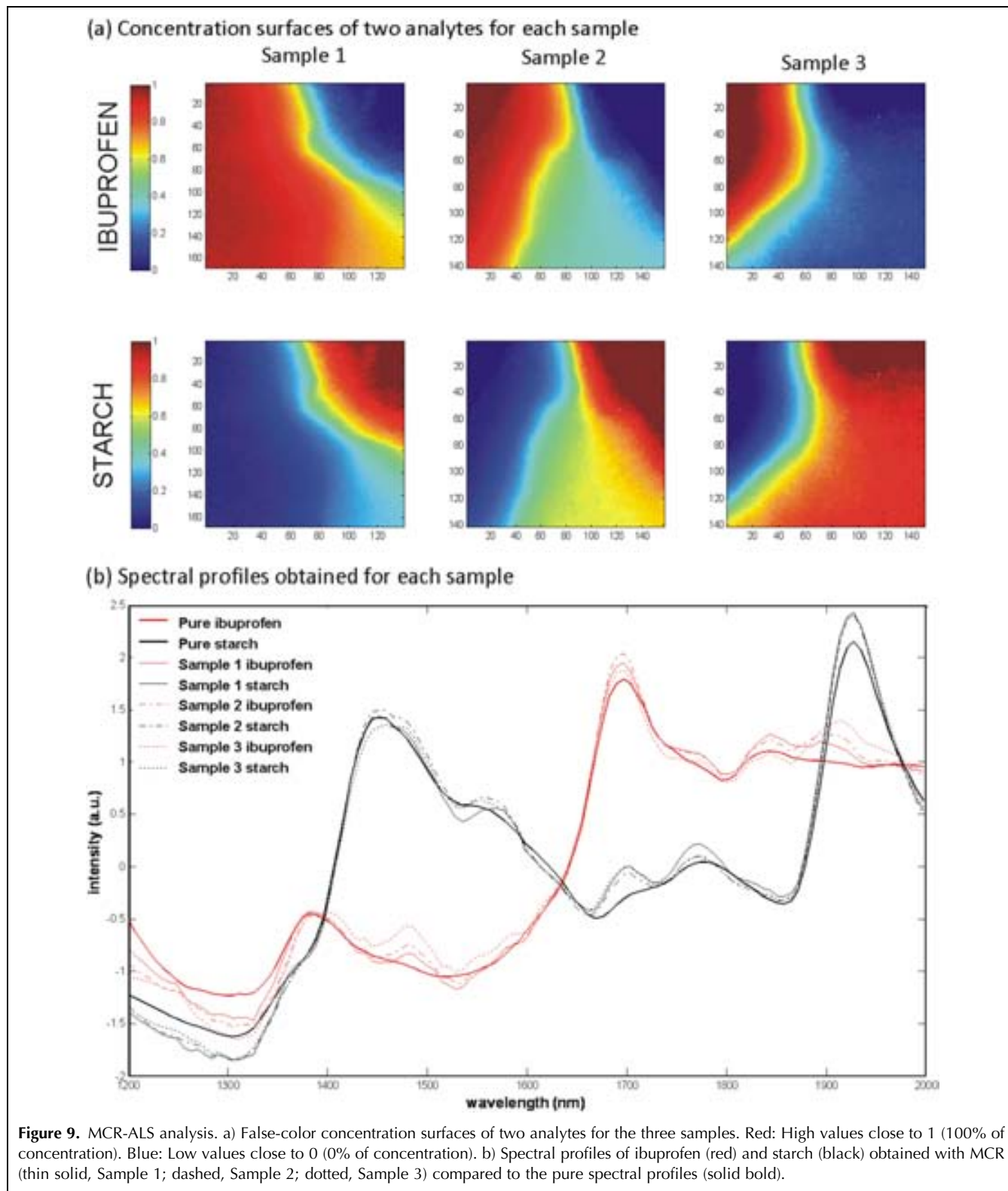
	Nominal concentration ^a		CLS concentrations ^a		MCR concentrations ^a		Augmented-MCR ^b	
	Ibuprofen	Starch	Ibuprofen	Starch	Ibuprofen	Starch	Ibuprofen	Starch
Sample 1	75	25	79	40	73	26	–	–
Sample 2	50	50	48	74	45	54	–	–
Sample 3	25	75	15	70	23	77	–	–
Sample 4	65	35	65	48	57	43	65	35

^aMean concentration (%) in the mixture area.^bMean concentration (%) in the Sample 4.

7.5. MCR-ALS and augmented-MCR-ALS results

The results of MCR-ALS for the three heterogeneous samples are shown in Fig. 9 and Table 2. Previous analysis of the number of components showed that the three samples could be explained by samples with two

components (results not shown). The models with two components explain more than 99% of the variance, and the areas of pure analytes and mixtures are perfectly defined (Fig. 9).



Since the closure constraint was imposed, it was expected that the concentration surfaces might indicate the real concentration of ibuprofen and starch. Table 2 shows the mean concentration calculated for each analyte in the mixture area. It can be seen that the concentration obtained for starch is slightly higher than the real one. This may be explained by the observation of the histograms of the correlation coefficients obtained with BACRA, where absorbance of starch is higher than ibuprofen. The comparison between the spectral profiles obtained with MCR-ALS and the pure spectral profiles is plotted in Fig. 9b. Their high correlations (upper than 0.95) indicate the goodness of the MCR-ALS models obtained.

For the homogeneous sample, the MCR-ALS model explained less than 90% of the total variance, and the mean concentrations obtained for ibuprofen and starch were 57% and 43%, respectively. The spectral profiles obtained in this analysis are plotted in red in Fig. 10, showing a very noisy shape and denoting the rank deficiency of the sample.

To improve the results, an augmented MCR-ALS model was considered by simultaneously analyzing the homogeneous sample and Sample 2 (full-rank image), as indicated in Fig. 2. The new results for the homogeneous sample are shown in Fig. 10 and Table 2. Now, the augmented MCR-ALS model explains more than 98% of the total variance. The mean values of concentration obtained for ibuprofen and starch (65% and 35%, respectively) perfectly match the theoretical ones. In

addition, the correlation between the spectral profiles obtained and pure components is above 0.980 for both analytes.

8. Conclusions and perspectives

We have shown the possibilities of several multivariate algorithms dealing with the study of NIR hyperspectral images involving a blending process of two analytes from a qualitative and quantitative point of view.

Fig. 11 summarizes the main benefits and the main drawbacks of each method when samples of a blending process were studied. This scheme reflects how studying blending samples with NIR-CI depends on the final purpose and on the availability of extra information (e.g., pure spectra of the components).

When only exploratory analysis is required, PCA, FCM clustering and BACRA have demonstrated possibilities. Of course, to get the pure spectra is very important in the decision of what methodology can be used. Nevertheless, this is not a real problem in pharmaceutical industry, where it is possible to obtain measurements for both components before blending them.

The combined use of PCA, FCM clustering and BACRA has demonstrated their usefulness, even in homogeneous samples, since the image can be affected by other sources of variability.

KM-cluster analysis could lead to misunderstanding or bad interpretation of results, since each pixel is exclu-

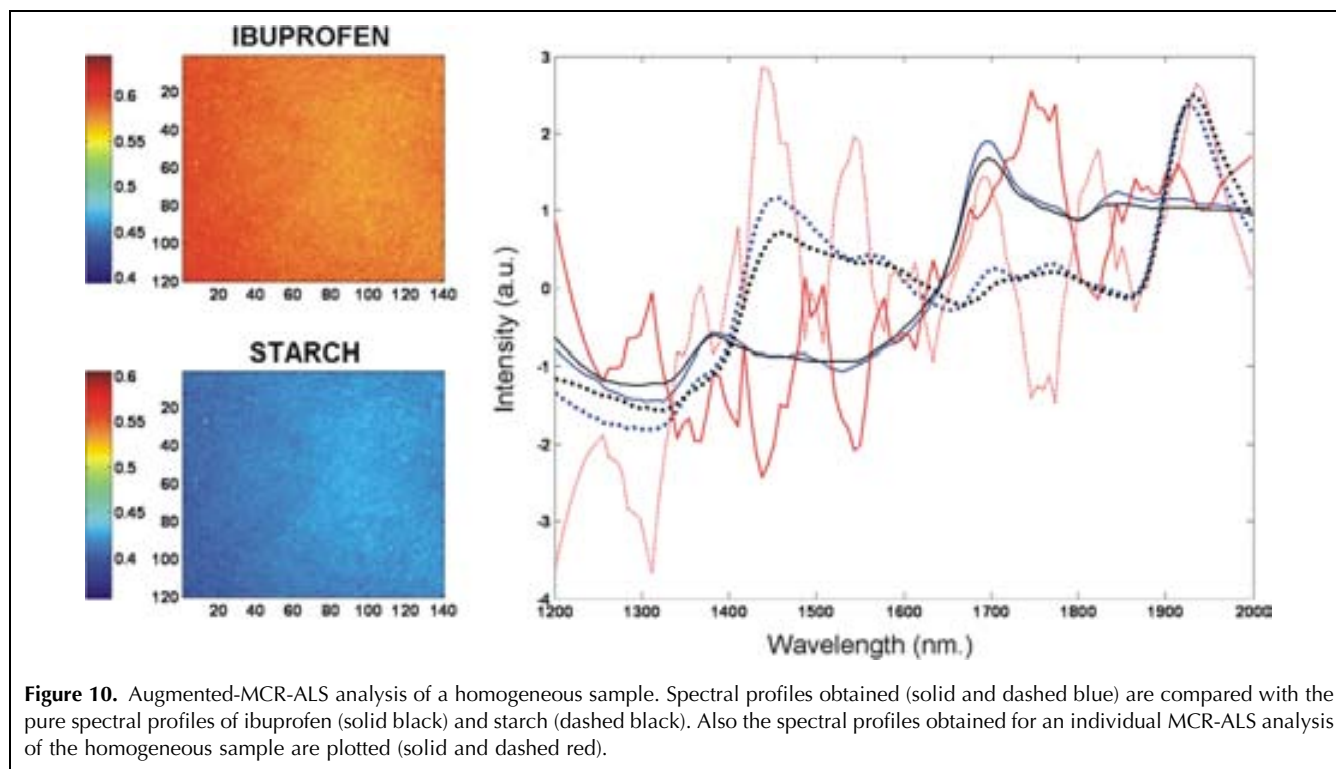
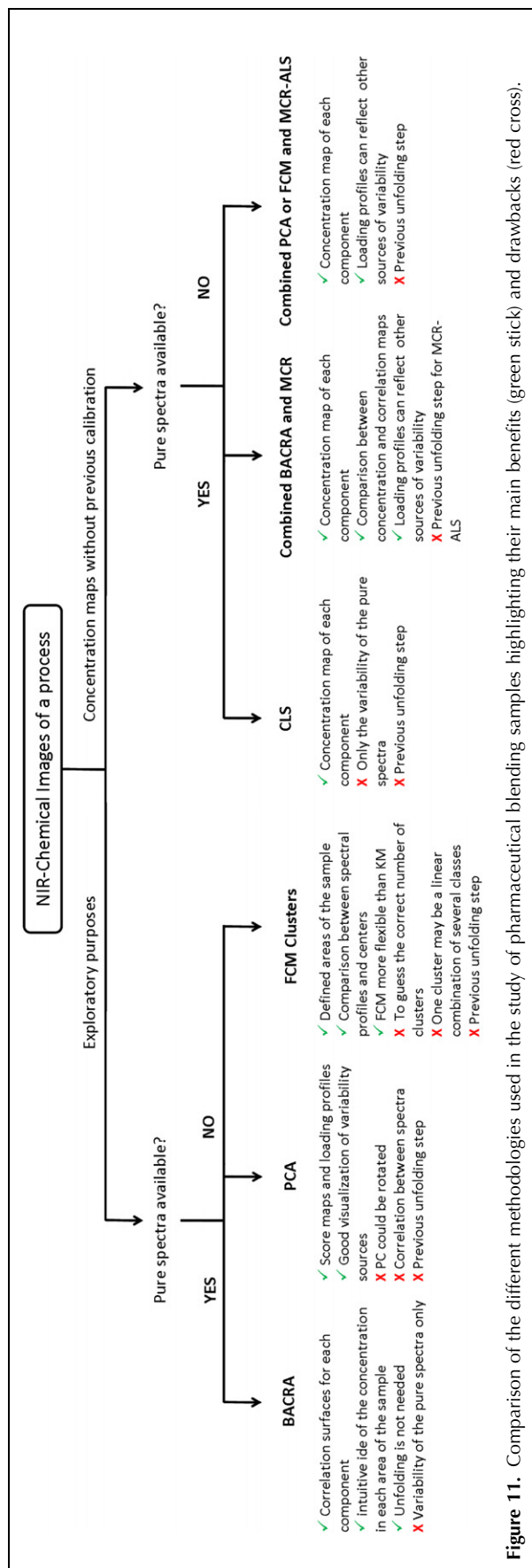


Figure 10. Augmented-MCR-ALS analysis of a homogeneous sample. Spectral profiles obtained (solid and dashed blue) are compared with the pure spectral profiles of ibuprofen (solid black) and starch (dashed black). Also the spectral profiles obtained for an individual MCR-ALS analysis of the homogeneous sample are plotted (solid and dashed red).



sively associated with one cluster. By contrast, FCM clustering offers the possibility of studying the surfaces but, instead of considering the number of clusters needed, the analysis has to be developed by considering the number of classes of the samples and the possible linear combination between classes that may occur in the mixture.

Correlation-coefficient mapping, developed with BACRA, is a good tool for studying the homogeneity of samples. Furthermore, mapping correlation-coefficient surfaces and histograms for each analyte pixel-by-pixel provides a general view of the distribution of analytes in the surface being studied. One drawback of correlation-coefficient mapping is that the pure spectra are needed.

For a more accurate study of the surfaces, two complementary techniques, CLS and MCR-ALS, can be very useful. CLS works perfectly if the spectra of the components are known and there is no other source of variability in the sample. The examples in this article have demonstrated that CLS does not work properly in predicting the concentration of ibuprofen and starch in the mixture area of samples. MCR-ALS using known spectra as initial estimations gives good results because MCR-ALS separates the pure signal from undesirable phenomena.

Of course, MCR-ALS always has the problem of the ambiguity associated with mathematical decomposition. Nevertheless, with a few proper constraints (non-negativity in concentration profiles and closure), high-quality results have been obtained. In addition, we showed that augmented MCR-ALS analysis is a powerful option for the analysis of real homogeneous samples, when the lack of selective areas presents a problem in MCR-ALS analysis

Acknowledgements

The authors want to thank to Projects CTQ2004-985 of Ministerio de Educación y Ciencia of Spain for the financial support. José Manuel Amigo wants to thank Frans Van Den Berg for his interest, comments and discussion about this work, and the Danish Research Council for Technology and Production Sciences for his post-doctoral fellowship.

References

- [1] E.N. Lewis, L.H. Kidder, E. Lee, K.S. Haber, *Spectrochemical Analysis Using Infrared Multichannel Detectors*, Blackwell Publishing Ltd., Oxford, UK, 2005.
- [2] V.W. Smail, A.K. Fritz, D.L. Wetzel, *Vib. Spectrosc.* 42 (2006) 215.
- [3] M. Politi, C.D. Tran, *J. Non-Cryst. Solids* 304 (2002) 64.
- [4] E. Lee, L.H. Kidder, V.F. Kalasinsky, J.W. Schoppelrei, E.N. Lewis, *Cytometry, Part A* 69A (2006) 888.
- [5] N. Jovanovic, A. Gerich, A. Bouchard, W. Jiskoot, *Pharm. Res.* 23 (2006) 2002.

- [6] T. Furukawa, H. Sato, H. Shinzawa, I. Noda, S. Ochiai, *Anal. Sci.* 23 (2007) 871.
- [7] J. Dubois, J.C. Wolff, J.K. Warrack, J. Schoppelrei, E.N. Lewis, *Spectroscopy (Amsterdam)* 22 (2007) 40.
- [8] S.G. Kazarian, K.W.T. Kong, M. Bajomo, J. Van der Weerd, K.L.A. Chan, *Food Bioprod. Process.* 83 (2005) 127.
- [9] K.L.A. Chan, N. Elkhider, S.G. Kazarian, *Chem. Eng. Res. Des.* 83 (2005) 1303.
- [10] K.L.A. Chan, S.G. Kazarian, *Lab Chip* 6 (2006) 864.
- [11] E. Lee, W.X. Huang, P. Chen, E.N. Lewis, R.V. Vivilecchia, *Spectroscopy (Amsterdam)* 21 (2006) 24.
- [12] S. Sasic, *Anal. Chim. Acta* 611 (2008) 73.
- [13] O.Y. Rodionova, L.P. Houmoller, A.L. Pomerantsev, P. Geladi, J. Burger, V.L. Dorofeyev, A.P. Arzamastsev, *Anal. Chim. Acta* 549 (2005) 151.
- [14] E.N. Lewis, J. Schoppelrei, E. Lee, L.H. Kidder, *Process Analytical Technology*, Blackwell Publishing, Oxford, UK, 2005.
- [15] E.N. Lewis, J. Schoppelrei, E. Lee, *Spectroscopy (Amsterdam)* 19 (2004) 26.
- [16] R.B. Shah, M.A. Tawakkul, M.A. Khan, *J. Pharm. Sci.* 96 (2007) 1356.
- [17] E.N. Lewis, L.H. Kidder, E. Lee, *Innovations Pharm. Technol.* (September 2006) 107.
- [18] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, *Trends Food Sci. Technol.* 18 (2007) 590.
- [19] G. Reich, *Adv. Drug Delivery Rev.* 57 (2005) 1109.
- [20] P. Geladi, H. Grahn, *Multivariate Image Analysis*, John Wiley & Son Ltd, New York, USA, 1997.
- [21] H. Grahn, P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis*, Wiley & Sons, Chichester, West Sussex, UK, 2007.
- [22] L. Zhang, M.J. Henson, S.S. Sekulic, *Anal. Chim. Acta* 545 (2005) 262.
- [23] A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, *Trends Anal. Chem.* 23 (2004) 70.
- [24] S. Chevallier, D. Bertrand, A. Kohler, P. Courcoux, *J. Chemometrics* 20 (2006) 221.
- [25] E.K. Kemsley, *Chemometrics Intelligent Lab. Syst.* 33 (1996) 47.
- [26] N. Elkhider, K.L.A. Chan, S.G. Kazarian, *J. Pharm. Sci.* 96 (2007) 351.
- [27] A.S. El-Hagrasy, H.R. Morris, F. D'Amico, R.A. Lodder, J.K. Drennen, *J. Pharm. Sci.* 90 (2001) 1298.
- [28] H. Ma, C.A. Anderson, *J. NIR Spectrosc. (Amsterdam)* 15 (2007) 137.
- [29] J. Burger, P. Geladi, *J. Chemometrics* 19 (2005) 355.
- [30] P. Geladi, J. Burger, T. Lestander, *Chemometrics Intelligent Lab. Syst.* 72 (2004) 209.
- [31] P. Geladi, B. Sethson, J. Nystrom, T. Lillhonga, T. Lestander, J. Burger, *Spectrochim. Acta, Part B* 59 (2004) 1347.
- [32] C.M. Andersen, R. Bro, *J. Chemometrics* 17 (2003) 200.
- [33] H.A.L. Kiers, J.M.F. Ten Berge, R. Bro, *J. Chemometrics* 13 (1999) 275.
- [34] R. Bro, C.A. Andersson, H.A.L. Kiers, *J. Chemometrics* 13 (1999) 295.
- [35] J. Burger, P. Geladi, *J. NIR Spectrosc.* 15 (2007) 29.
- [36] R.J. Barnes, M.S. Dhanoa, S.J. Lister, *Appl. Spectrosc.* 43 (1989) 772.
- [37] P. Geladi, D. Macdougall, H. Martens, *Appl. Spectrosc.* 39 (1985) 491.
- [38] W. Windig, *Chemometrics Intelligent Lab. Syst.* 23 (1994) 71.
- [39] L. Zhang, *Pattern Recognition Methods for the Analysis of Infrared Imaging Data and Multivariate Calibration Standardization for Near Infrared Spectroscopy*, PhD Thesis, Ohio University, USA, 2002.
- [40] O. Svensson, K. Abrahamsson, J. Engelbrektsson, M. Nicholas, H. Wikstrom, M. Josefson, *Chemometrics Intelligent Lab. Syst.* 84 (2006) 3.
- [41] M. Maeder, A. Zilian, *Chemometrics Intelligent Lab. Syst.* 3 (1988) 205.
- [42] B.M. Wise, N.B. Gallagher, S.W. Butler, D.D. White, G.G. Barna, *J. Chemometrics* 13 (1999) 379.
- [43] P. Geladi, H. Isaksson, L. Lindqvist, S. Wold, K. Esbensen, *Chemometrics Intelligent Lab. Syst.* 5 (1989) 209.
- [44] P. Geladi, S. Wold, K. Esbensen, *Anal. Chim. Acta* 191 (1986) 473.
- [45] S. Sasic, *Appl. Spectrosc.* 61 (2007) 239.
- [46] Y. Roggo, A. Edmond, P. Chalus, M. Ulmschneider, *Anal. Chim. Acta* 535 (2005) 79.
- [47] Y. Roggo, N. Jent, A. Edmond, P. Chalus, M. Ulmschneider, *Eur. J. Pharm. Biopharm.* 61 (2005) 100.
- [48] F. Clarke, *Vib. Spectrosc.* 34 (2004) 25.
- [49] B. Walczak, *Wavelets in Chemistry*, Elsevier, New York, USA, 2000.
- [50] S. Theodoridis, K. Koutroumbas, *Pattern Recognition*, Academic Press, New York, USA, 1999.
- [51] J.C. Noordam, W.H.A.M. van den Broek, P. Geladi, L.M.C. Buydens, *Chemometrics Intelligent Lab. Syst.* 75 (2005) 115.
- [52] J.B. MacQueen, *Some Methods for Classification and Analysis of Multivariate Observations*, University of California Press, Berkeley, CA, USA, 1967.
- [53] Y.J. Zhang, W.N. Wang, X.N. Zhang, Y. Li, *Inf. Sci.* 178 (2008) 1205.
- [54] L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, Chichester, West Sussex, UK, 1990.
- [55] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*, Plenum Press, New York, USA, 1981.
- [56] S.J. Hamilton, A.E. Lowell, R.A. Lodder, *J. Biomed. Opt.* 7 (2002) 561.
- [57] T.T. Lied, P. Geladi, K.H. Esbensen, *J. Chemometrics* 14 (2000) 585.
- [58] J. Burger, P. Geladi, *Analyst (Cambridge UK)* 131 (2006) 1152.
- [59] C. Gendrin, Y. Roggo, C. Collet, *Talanta* 73 (2007) 733.
- [60] W.Y. Li, A. Woldu, R. Kelly, J. McCool, R. Bruce, H. Rasmussen, J. Cunningham, D. Winstead, *Int. J. Pharmaceutics* 350 (2008) 369.
- [61] R. Tauler, *Chemometrics Intelligent Lab. Syst.* 30 (1995) 133.
- [62] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, *Chemometrics Intelligent Lab. Syst.* 76 (2005) 101.
- [63] A. de Juan, R. Tauler, *Crit. Rev. Anal. Chem.* 36 (2006) 163.
- [64] http://www.models.life.ku.dk/users/jose_manuel_amigo/index.htm.
- [65] <http://www.ub.es/gesq/mcr.htm>.

7. CONCLUSIONES

Debido a que se han abordado estudios diferentes, las conclusiones se exponen a continuación en función del estudio realizado:

1) Estudios sobre la determinación de API en preparados farmacéuticos.

- Se demuestra la capacidad de la Espectroscopia NIR en la determinación de analitos minoritarios (~0.02%) en muestras de geles hidroalcohólicos, que constituyen una matriz altamente absorbente. Este logro representa superar ampliamente la sensibilidad que la bibliografía da para la Espectroscopia NIR, haciendo posible determinaciones a bajas concentraciones incluso en matrices absorbentes. Un correcto diseño y preparación de las muestras de calibración ha permitido obtener resultados analíticos exactos de analitos minoritarios. La aplicación de diseños experimentales adecuados disminuye las correlaciones indeseadas entre el analito y el resto de componentes, lo que da lugar a modelos robustos.
- Se ha desarrollado una estrategia que permite la determinación del contenido de API durante el proceso de fabricación: durante un intermedio del proceso y al final del mismo. El primero tiene como objetivo asegurar la correcta fabricación durante el proceso y el segundo la confirmación del contenido en el producto acabado. Estos modelos sustituyen con ventaja a los métodos convencionales de determinación de API con la consiguiente ventaja en tiempo y en coste.
- Los modelos desarrollados se han preparado usando únicamente muestras de laboratorio; esta estrategia libera a la espectroscopia NIR de la exigencia de aplicación de métodos de referencia para analizar las

CONCLUSIONES

- muestras que se utilizan en la construcción de los modelos y a la vez simplifica la preparación de las muestras de calibración.
- Ha sido posible descartar de forma inequívoca la transformación polimórfica de un API durante el proceso de fabricación; una publicación había recogido esta transformación en un proceso de granulación húmeda con celulosa microcristalina como excipiente mayoritario. La granulación con sacarosa como excipiente no conduce a la transformación polimórfica. Este estudio ha implicado la utilización de algoritmos de MCR – ALS y ha permitido descartar la transformación por no tener lugar cambios en el espectro del API (idénticos espectros antes y después de la granulación)

2) Monitorización de un proceso de granulación húmeda mediante medidas no invasivas.

- Se ha desarrollado una nueva metodología para el seguimiento espectroscópico de un proceso de granulación húmeda en una granuladora de lecho fluido, como un primer estudio de su aplicación a PAT. Los espectros obtenidos han permitido realizar un estudio tanto cualitativo como cuantitativo de las 3 etapas del proceso.
- Un análisis multivariante de los espectros NIR mediante un Análisis por Componentes Principales, permite diferenciar las tres etapas del proceso. Los dos primeros Componentes Principales explican casi toda la variabilidad de las muestras del proceso y se han identificado las 3 etapas del proceso mediante el gráfico de scores v.s. tiempo. La interpretación de los loadings espectrales permite asociar las fuentes

CONCLUSIONES

- de variabilidad: el primer componente principal está relacionado con la evolución del tamaño de partícula a lo largo del proceso, mientras que el segundo componente principal explica la variación del contenido de agua.
- Los modelos de calibración PLS construidos permiten realizar la determinación de los parámetros pertinentes: determinación del contenido de humedad, fracción de muestra entre tamaños de partícula establecidos y densidad, que permiten seguir estos parámetros en tiempo real.

3) Estudio de técnicas de Imagen Química en el campo farmacéutico.

- El estudio realizado se ha orientado a la prospección de las técnicas quimiométricas más apropiadas para extraer la información química de interés de la información espectral contenida en una imagen. En el estudio se han discutido las posibilidades de las diferentes técnicas ensayadas y su aplicación al campo farmacéutico, principalmente en el tema de homogeneidad de distribución del API en preparaciones farmacéuticas.
- La aplicación de una técnica muy simple como es el coeficiente de correlación permite abordar este estudio de forma rápida y visualizar esta distribución a través de los mapas de correlación.
- Diferentes técnicas de calibración pueden aportar soluciones a la determinación cuantitativa, pero destaca la MCR-ALS por los escasos requisitos necesarios para obtener la información cuantitativa (únicamente los espectros de los componentes, que constituyen una información muy asequible en este campo)

CONCLUSIONES

Aumentando la matriz con espectros puros mejora sustancialmente los resultados finales.

- La implementación del coeficiente de correlación, asociado a otras herramientas en un software (BACRA), desarrollado dentro del grupo de investigación, constituye una interesante aportación para el estudio de imágenes hiperespectrales.

7. CONCLUSIONS

Because different studies have been proposed, the conclusions are presented below according to the study:

1) Studies on the determination of API in pharmaceutical preparations.

- It demonstrates the ability of NIR spectroscopy in the determination of low concentration analytes (~ 0.02%) in samples of hydro-alcoholic gels, which show a highly absorbing matrix. These results show that the sensitivity that far exceeds the literature given to NIR spectroscopy, enabling determinations at low concentrations even in absorbing matrices. A proper design and preparation of the calibration samples has yielded good analytical results of low concentration analytes. The application of appropriate experimental designs reduces unwanted correlations between the analyte and other components, leading to robust models.
- We have developed a strategy that allows the determination of API during the manufacturing process: during an intermediate step and the end of it. The first aims to ensure proper manufacturing process and during the second confirmation of the API content in the finished product. These models replaced with advantage to the conventional methods for determining API with the consequent advantage in time and cost.
- The models developed have been prepared using only laboratory samples this strategy relieves the NIR spectroscopy of the requirement for implementation of reference methods for analyzing samples that are used in the construction of models and also simplifies the preparation of calibration samples.

CONCLUSIONS

- It has been possible to rule out unequivocally the polymorphic transformation of an API during the manufacturing process, a publication had picked this transformation process in a wet granulation with microcrystalline cellulose as the main excipient. Granulation with sucrose as an excipient does not lead to the polymorphic transformation. This study has involved the use of algorithms MCR - ALS and has helped out the transformation take place not by changes in the spectrum of the API (identical spectra before and after granulation)

2) Monitoring of a wet granulation process using non-invasive measures.

- We have developed a new spectroscopic method for monitoring a process of wet granulation in a fluid bed granulator such as a first study of its application to TAP. The spectra obtained have allowed a qualitative and quantitative study of the 3 stages.
- A multivariate analysis of NIR spectra using a Principal Component Analysis allows distinguishing three stages. The two first principal components explain almost all the variability of samples of the process and have identified 3 stages of the process through graphic scores vs time. The interpretation of spectral loadings allows relating the sources of variability: the first principal component is related to the evolution of particle size throughout the process, while the second principal component explains the variation of water content.
- PLS calibration models constructed allow the identification of relevant parameters: determination of moisture content between sample particle size and density set to allow for these parameters in real time.

3) Study of imaging techniques in pharmaceutical chemistry.

CONCLUSIONS

- This study has been focussed to the exploration of the most appropriate chemometric techniques to extract the chemical information of interest to the spectral information contained in an image. In the study we discussed the possibilities of the different techniques tested and its application in the pharmaceutical field, mainly on the issue of uniformity of distribution of the API in pharmaceutical preparations.
- The implementation of a very ease technique as the correlation coefficient allows a quick study and visualizes this distribution through the maps of correlation.
- Different calibration techniques can provide solutions to the determination but emphasizes the MCR-ALS for scarce requirements to obtain the quantitative information (only the spectra of components are required, which are a very affordable in this area). Increasing the matrix with pure spectra substantially improved outcomes.

The implementation of the correlation coefficient, associated with other tools in a software (Bacra) developed within the research group, is an interesting contribution to the study of hyperspectral images.

