

UNIVERSIDAD AUTÓNOMA DE BARCELONA

DEPARTAMENTO DE QUÍMICA

ESTUDIO DE LA INFLUENCIA DE PARÁMETROS  
FÍSICOS EN LA CONSTRUCCIÓN DE MODELOS  
DE CALIBRACIÓN EN LA ESPECTROSCOPIA EN  
EL INFRARROJO CERCANO

AUTOR: DÁMARIH VALDÉS CALVIÑO

DIRECTOR: DR. MARCELO BLANCO ROMIA

BELLATERRA, 2005

*A mis padres,  
eterna fuente de inspiración.*

*“Nada sugiere tanta y tan hermosa literatura como un párrafo de ciencia.”*

*José Martí*

## AGRADECIMIENTOS

---

*Haber terminado este trabajo ha sido muy gratificante y ha estado lleno de momentos reconfortantes, pero también de momentos difíciles. A lo largo de esta travesía he contado con el apoyo de muchísimas personas a las cuales les quiero expresar mi gratitud, pues sin ellas seguramente este barco no habría llegado exitosamente a tierra firme. Expresar los sentimientos con palabras es una tarea difícil, pero me aventuro a ello.*

*Al Dr. Marcelo Blanco por haberme dado la oportunidad de iniciar este trabajo y haberme abierto las puertas del maravilloso universo de la espectroscopia NIR. Igualmente quiero agradecer a los profesores Hortensia Iturriaga, Jordi Coello y Santiago MasPOCH por sus aportes científicos a lo largo de estos años.*

*Concluir este trabajo ha sido también posible gracias a mis profesores Dr. Miguel Catasús y Dra. Roser Rubio, ambos han sido un eslabón importante en mi formación profesional. Admiro en ellos su gran nivel científico y su excelente calidad humana, que hacen posible que me sienta cerca de ellos a pesar del paso del tiempo.*

*Agradezco a mis compañeros del Grupo de Quimiometría Aplicada (a los pasados y a los presente; a los catalanes, castellanos, andaluces y de toda la península ibérica) pues hemos compartido muchas horas en el laboratorio y con cada uno he aprendido algo. A Manel quiero agradecerle de modo especial, no solamente por sus contribuciones científicas a lo largo de esta etapa, sino también por su maravillosa amistad que me ha abierto las puertas de Cataluña y ha hecho posible que me sienta como en casa. Gracias también a su familia, pues tener el amor de una familia cuando se esta lejos de casa, es un valiosísimo tesoro.*

*Al personal de Secretaria de Química, muy especialmente a Cecilia Inchausti, Puri Motta y Angels Solvas quiero agradecerles por su eficiente y humana gestión administrativa. Gracias también a José Manuel Catalán un buen laborante con un especial sentido del humor.*

*No podría dejar de mencionar a mis adorados amigos, que son la familia que la naturaleza no me dio, pero afortunadamente gané en el camino de la vida. A Giselle, la de los años, la de siempre... a Yencita, por su ternura y su paz... a Pio, por seguir siendo mi eterno amigo de la Universidad... a Yován, por sus exquisitas comiditas y por algún que otro “craqueo” al estilo pinareño... a Xavi, por compartir esa gran pasión hacia Barcino... a*

## AGRADECIMIENTOS

---

*Wilfre, por su poder mediático en aras de la comprensión y la conservación de la amistad... a Suany y Tamo, por mantener el cuarteto del bachillerato... a Osvi, Julio, Carlos E., Dacal, Daniel, etc. por las noches inolvidables en nuestra amada Habana... a F. Juncal, porque siempre, siempre, estás ahí... a Lisy y la peña médica de Manrusia, por su gran calidad médica y humana. Por favor el orden de los factores no altera el producto, para cada uno hay un espacio inmenso dentro de mi corazón!!!*

*Quiero agradecer especialmente a mi familia de Santa Fé, Nuevo Vedado, Vedado y 502 que si bien estamos separados por un inmenso océano, nos mantenemos unidos por los lazos indestructibles del amor. También quiero expresar mi gratitud hacia la familia de Gastón -que como muchos andamos desperdigados por el planeta- en la Florida, North Carolina y “Le Danseuse”.*

*Un agradecimiento muy especial para mis abuelas Abue y Tatica, seres maravillosos en la vida de un ser humano, que tanta dulzura y protección me han brindaron desde la cuna. Tatica, donde quiera que estés, sabrás que esto ha sido posible también a tu ejemplo de disciplina y perseverancia.*

*A mis hermanas Yai, Diani y Laura, quiero repetirles cuanto las quiero y cuan importante son para mi cada día, a pesar de ese maldito océano. Por ustedes intento dar lo mejor de mi cada día.*

*Engendrar a un ser humano es un acto maravilloso y a su vez difícil, pero más aun es guiarlo por el arduo camino de la vida. Mami y papi, creo sinceramente que lo han hecho de forma excepcional y me faltan palabras para agradecerles todo lo que han hecho y siguen haciendo por mí. La culminación de esta tesis, es también una obra de ustedes.*

*Por ultimo quiero agradecer de modo especial a Gastón por su colaboración y fe en este proyecto, acompañado de buenas dosis de comprensión, paciencia e infinito amor. Empezar y terminar este proyecto juntos, ha sido un sueño para ambos y una motivación para lanzarnos en la arriesgada pero gratificante aventura de ser padres.*

*A todos, de corazón, mi eterna gratitud.*

*Dámarih*

**ABSTRACT**

**OBJETIVOS**..... 1

**INTRODUCCION**

|   |    |
|---|----|
| 1. Espectroscopia en el Infrarrojo Cercano .....                                    | 3  |
| 1.1 Breve reseña histórica.....   | 3  |
| 1.2 Fundamentos teóricos.....   | 4  |
| 1.3 Instrumentación .....   | 10 |
| 1.3.1 Fuente de radiación.....  | 11 |
| 1.3.2 Sistema de selección de longitud de onda .....                                | 12 |
| 1.3.3 Compartimiento de la muestra .....  | 13 |
| 1.3.4 Detectores.....   | 16 |
| 1.4 Aplicaciones de la Espectroscopia en el Infrarrojo Cercano.....                 | 16 |
| 2. Quimiometría .....   | 23 |
| 2.1 Análisis Cualitativo.....   | 23 |
| 2.1.1 Métodos de Reconocimiento de Pautas (PRMs) .....                              | 24 |
| 2.1.1.1 Métodos No Supervisados .....   | 26 |
| 2.1.1.2 Métodos Supervisados .....  | 30 |
| 2.1.1.3 Bibliotecas de Identificación .....   | 34 |
| 2.2 Análisis Cuantitativo.....  | 35 |
| 2.2.1 Regresión Lineal Múltiple (MLR).....  | 36 |
| 2.2.2 Regresión en Componentes Principales (PCR) .....                              | 37 |
| 2.2.3 Regresión Parcial por Mínimos Cuadrados (PLS).....                            | 39 |
| 2.2.4 Estrategias generales en el análisis.....                                     | 39 |
| 2.2.5 Validación del método de análisis .....                                       | 43 |
| 3. Polimorfismo .....   | 51 |
| 3.1 Definición y propiedades .....  | 51 |
| 3.2 <i>Pseudo</i> -polimorfos. Formas amorfas, solvatos e hidratos.....             | 52 |
| 3.3 Métodos de caracterización .....  | 53 |
| 3.4 Efecto del procesamiento farmacéutico en fármacos polimórficos y solvatos ..... | 55 |

**METODOLOGIA Y DISCUSION DE LOS RESULTADOS**

|  |    |
|--|----|
| 4. Introducción .....  | 61 |
| 5. Metodología .....   | 62 |
| 5.1 Descripción y preparación de las muestras.....   | 62 |
| 5.2 Instrumentación.....   | 63 |
| 5.3 <i>Software</i> .....  | 65 |
| 5.4 Procesamiento de datos .....   | 66 |
| 6. Discusión de los Resultados .....   | 71 |
| 6.1 Caracterización y análisis de formas polimórficas.....   | 71 |
| 6.1.1 Introducción .....   | 71 |
| 6.1.2 Caracterización de <i>pseudo</i> -polimorfos por difracción de Rayos X .....                     | 73 |
| 6.1.3 Análisis de <i>pseudo</i> -polimorfos mediante espectroscopia en el infrarrojo cercano .....     | 75 |
| 6.1.3.1 Modelos de calibración (PLS).....  | 76 |
| 6.1.3.2 Estudio de Estabilidad .....   | 81 |
| 6.2 Estudio y corrección del efecto de la temperatura en espectroscopia en el infrarrojo cercano ..... | 85 |
| 6.2.1 Introducción .....   | 85 |
| 6.2.2 Estudio del efecto de la temperatura sobre los espectros NIR .....                               | 89 |
| 6.2.3 Estrategias para la corrección del efecto de la temperatura.....                                 | 92 |
| 6.2.3.1 Modelar el efecto de la temperatura .....  | 92 |
| 6.2.3.2 Eliminar el efecto de la temperatura .....   | 95 |

|                           |     |
|---------------------------|-----|
| <b>CONCLUSIONES</b> ..... | 105 |
|---------------------------|-----|

**ANEXOS**

Anexo I: "Characterization and analysis of polymorphs by near-infrared spectrometry"

Anexo II: "Application of NIR Spectroscopy in Polymorphic Analysis: Study of *Pseudo*-Polymorphs Stability"

Anexo III: "Influence of temperature on the predictive ability of near infrared spectroscopy models"

Anexo IV: "Suppressing the temperature effect in Near Infrared Spectroscopy by using Orthogonal Signal Correction"

## ABSTRACT

---

In the last years there has been an increase in the use of near infrared spectroscopy, due to its multiple advantages: it is a non-invasive and non-destructive technique, it requires very little or no sample preparation, the ability to provide measurements and results in a highly expeditious manner, the low cost of analysis, an absence of reagents; as well as the possibility of making simultaneous determination of physical and chemical parameters.

Nonetheless, the characteristic of the NIR signal requires the application of multivariate methods to obtain the desired information contained in the spectrum. The spectral variability produced by the insufficient control of physical parameters, gives rise to complex models of calibration as well as a poorer predictive capacity.

In an industrial process experimental conditions can not always be strictly controlled and variations of the physical conditions of the system (temperature, pressure, moisture, etc.) often take place. The different arrangement from molecules in a crystalline structure gives rise to the polymorphism and this property complicate the models of calibration of samples in which this property is present.

In the pharmaceutical industry is very common that active principles can exist in two or more different crystalline forms (polymorphs). We have taken advantage of these differences for the characterization and determination of different polymorphs of a substance. One of the objectives of this thesis is to develop new methodologies of analysis for the identification and characterization of polymorphs of azithromycin by means of NIR spectroscopy. Models of calibration were developed based on the algorithm of partial least-square (PLS) regression for the determination of a polymorphic form in the presence of another and were validated for their application in the

## ABSTRACT

---

pharmaceutical industry. In addition, a study of stability under high temperature and moisture conditions was carried out to establish the stability of different forms.

Temperature changes alter the intensity of absorption bands and shift their position in the NIR spectra for liquid samples. The second objective is to study the effect of temperature on NIR spectra on the ingredients of an esterification reaction over the range 25-90 °C and to construct calibration models to determine the concentration of the components of the mixtures within the temperature range with a correct predictive capacity. Two strategies were used; one consists of modeling and the other in eliminating the temperature effect. In order to model the temperature effect calibration models were constructed, that implicitly include the variable temperature in different ranges, using the partial least-square (PLS) regression and stepwise principal component regression (MLR stepwise). Another strategy is to eliminate the temperature effect by means of the variable filtering method known as orthogonal signal correction (OSC). Once the spurious spectral information that is unrelated to the concentration of the components of the mixtures was removed, models of calibration are constructed based on the partial least-square (PLS) regression.

In conclusion, this thesis provides analytical tools of interest applicable to the pharmaceutical industry for the characterization of polymorphs of azithromycin and the determination of polymorphs mixtures as well as providing a suitable method for the correction of the temperature effect in liquid samples.

## Objetivos

El desarrollo de nuevas metodologías de análisis simples, rápidas y fiables es una de las demandas más frecuentes de la industria. En los últimos años se ha producido un auge en el empleo de la espectroscopia en el infrarrojo cercano, debido a sus múltiples ventajas como son: rapidez y simplicidad en la obtención de los espectros y de los resultados del análisis, posibilidad de analizar muestras líquidas y sólidas sin necesidad de pretratamiento alguno por lo que no se generan residuos, instrumentación versátil y adaptable a las necesidades de las medidas (off-, at-, on-, in-line) en la industria, así como la posibilidad de determinar simultáneamente parámetros químicos y físicos. Sin embargo, la baja selectividad y que el espectro sea una señal compleja, hace necesario el uso de técnicas multivariantes para extraer la información de interés analítico.

En la presente memoria se estudia la influencia de parámetros físicos en el desarrollo de modelos de calibración, que si bien permiten determinar parámetros físicos, también complican los modelos para determinar la composición química. Se han estudiado dos problemáticas diferentes:

1. Efecto de la forma de cristalización del analito que aumenta la complejidad del modelo de determinación del mismo, pero que también permite el desarrollo de metodologías de análisis para la identificación y caracterización de distintas formas polimórficas.
2. Desarrollo de metodologías para la corrección del efecto que la temperatura ejerce sobre los espectros NIR. Se estudiarán diferentes estrategias para obtener modelos de calibración con una correcta capacidad predictiva en todo el rango de temperatura en estudio.



# INTRODUCCION

# Capítulo 1

## 1. Espectroscopia en el Infrarrojo Cercano

1.1 Breve reseña histórica

1.2 Fundamentos teóricos

1.3 Instrumentación

1.3.1 Fuente de radiación

1.3.2 Sistema de selección de longitud de onda

1.3.3 Compartimiento de la muestra

1.3.4 Detectores

1.4 Aplicaciones de la Espectroscopia en el Infrarrojo Cercano

## 1. Espectroscopia en el Infrarrojo Cercano

### 1.1 Breve reseña histórica

Friedrich Wilhelm Herschel en el año 1800 observó que existía energía después del rojo y como él no podía observarla, le nombró *infrared (IR)*, usando el prefijo latino *infra-* que significa antes [1].

Durante el siglo XIX se hicieron escasos progresos en este campo. En 1881 W. Abney y E.R. Festing registraron el espectro de líquidos orgánicos en el rango de 1-1.2  $\mu\text{m}$  [2]. Este trabajo tuvo una especial relevancia pues constituyó la primera medida en el infrarrojo cercano y también las primeras interpretaciones. Desafortunadamente, la carencia de equipos de detección impedía hacer progresos reales fuera de la región del visible.

En los inicios del siglo XX, la naturaleza del espectro electromagnético ya era mucho mejor comprendida. En 1905 William Weber Coblentz, motivado por el trabajo de W. Abney y E.R. Festing, construyó un espectrómetro empleando un prisma de sal de roca y una termopila conectada a un galvanómetro de espejo [3].

Durante la Segunda Guerra Mundial, se impulsó el desarrollo de semiconductores de PbS como detectores infrarrojos para la determinación de calor. En la década de los 1950's finalmente la espectrometría en el infrarrojo cercano, contaba con detectores de buena sensibilidad [4]. Si bien la primera mitad del siglo XX fue escasa en publicaciones, a partir de la década de los 50's se produce un importante auge de las investigaciones en el infrarrojo cercano.

Finalmente, un aspecto de gran relevancia en la evolución de la espectroscopia en el infrarrojo cercano (NIRS) ha sido el desarrollo de los ordenadores, no solamente por la facilidad para controlar y adquirir la información del instrumento, sino también para facilitar las calibraciones y el análisis de datos.

## 1.2 Fundamentos teóricos

La región del infrarrojo comprende las longitudes de onda de 780 a  $10^6$  nm y se divide en tres partes como se muestra en la **Figura 1**.

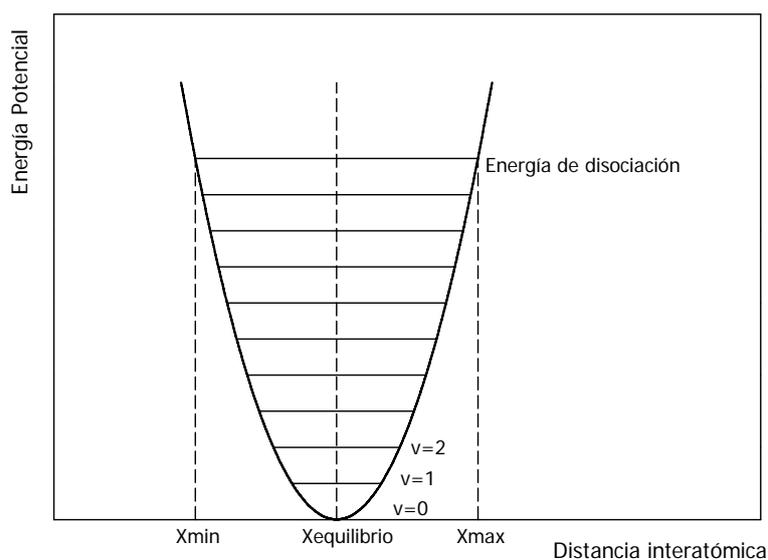
| Infrarrojo cercano<br>(NIR) | Infrarrojo medio<br>(MIR o IR) | Infrarrojo lejano<br>(FIR) |
|-----------------------------|--------------------------------|----------------------------|
| 780-2500 nm                 | 2500-40000 nm                  | $4 \cdot 10^4$ - $10^6$ nm |

**Fig. 1** Regiones de la región del infrarrojo.

Para absorber radiación infrarroja, una molécula debe experimentar un cambio neto en el momento dipolar como consecuencia de los movimientos de vibración o de rotación. La fluctuación de la densidad electrónica entorno de la molécula crea un campo eléctrico que puede interactuar con el campo eléctrico asociado a la radiación. Si la frecuencia eléctrica de la radiación iguala la frecuencia natural de vibración de la molécula, se produce un cambio en la amplitud de la vibración de la molécula y como consecuencia se absorbe la radiación.

Las moléculas diatómicas homonucleares ( $H_2$ ,  $O_2$ ,  $N_2$ ,  $Cl_2$ , etc.) no absorben en la región del infrarrojo, ya que no hay cambio en el momento dipolar de estas moléculas.

La vibración de una molécula diatómica puede aproximarse a la de un **oscilador armónico**. Cuando se produce la vibración, la energía potencial cambia continuamente, acercándose a cero cuando los átomos se acercan a la posición de equilibrio y haciéndose máxima cuando la distancia entre los átomos es mínima o máxima (**Figura 2**).



**Fig. 2** Energía Potencial en el modelo del oscilador armónico.

La energía potencial en el oscilador armónico se define como:

$$E = \frac{1}{2} kx^2 \quad (1.1)$$

donde  $k$  es la constante de la fuerza y  $x$  el desplazamiento de los átomos.

La frecuencia natural de vibración de un sistema constituido por dos masas unidas por un muelle perfectamente elástico se puede expresar mediante la Ley de Hooke:

$$\nu = \frac{1}{2\pi} \left( \frac{k(m_1 + m_2)}{m_1 m_2} \right)^{1/2} \quad (1.2)$$

donde  $\nu$  es la frecuencia natural de la vibración,  $k$  la constante de la fuerza,  $m_1$  y  $m_2$  la masa de cada átomo.

Sin embargo, los postulados de la mecánica cuántica plantean que la energía de vibración de dos átomos es discontinua y las transiciones solo pueden ocurrir entre niveles discretos de energía. Por tanto las ecuaciones descritas anteriormente no describen completamente el comportamiento de las partículas atómicas.

La energía vibracional, resolviendo ecuaciones cuánticas se pueden expresar como:

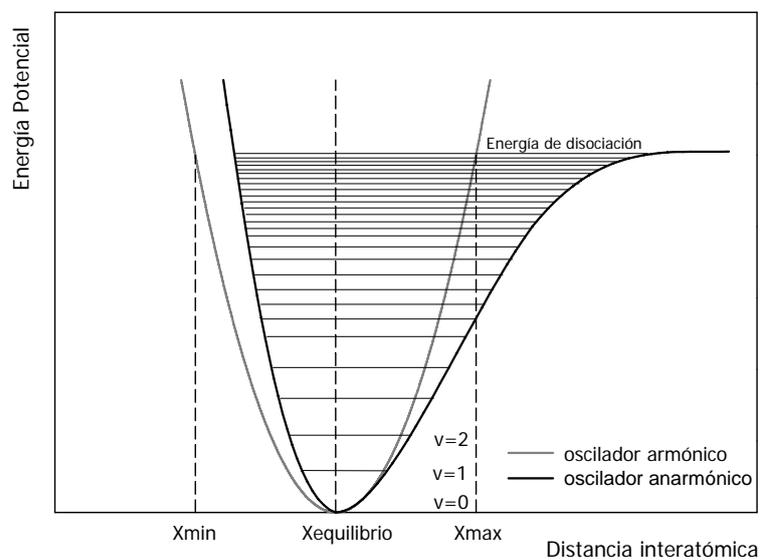
$$E_{vib} = \left( \nu + \frac{1}{2} \right) h \nu \quad (1.3)$$

donde  $\nu$  es el número cuántico vibracional (que solo puede tener valores enteros),  $h$  la constante de Planck y  $\nu$  la frecuencia de la vibración.

Combinando las ecuaciones 1.2 y 1.3 se obtiene:

$$E_{vib} = \left( \nu + \frac{1}{2} \right) \frac{h}{2\pi} \left( \frac{k(m_1 + m_2)}{m_1 m_2} \right)^{1/2} \quad (1.4)$$

Este modelo del oscilador armónico que tiene en cuenta consideraciones cuánticas, resulta incompleto para describir moléculas reales. Cuando dos átomos se acercan aparece una repulsión coulombica entre los dos núcleos pero en sentido contrario del movimiento, lo que implica un aumento más rápido de la energía potencial. Por otra parte, cuando los átomos se alejan, la distancia interatómica se acerca a la distancia de rotura del enlace, con la consiguiente disminución de la energía potencial. Las moléculas reales se acercan más al comportamiento de un **oscilador anarmónico (Figura 3)**. Las curvas armónicas y anarmónicas son muy similares a energías bajas, por lo que las moléculas entorno a la posición de equilibrio tienen un comportamiento armónico.



**Fig. 3** Energía Potencial en el modelo del oscilador anarmónico.

La energía vibracional corregida con términos que explican la anarmonicidad de las moléculas se define como:

$$E_{vib} = \left(\nu + \frac{1}{2}\right)h\nu - \left(\nu + \frac{1}{2}\right)^2 h\nu y - \left(\nu + \frac{1}{2}\right)^3 h\nu y' - \dots \quad (1.5)$$

donde  $y, y', \dots$  son las constantes de anarmonicidad. Para valores pequeños de  $\nu$ , el tercer término y siguientes pueden ser ignorados.

En el modelo anarmónico los niveles de energía no están igualmente espaciados. A niveles de energía más altos, la energía entre niveles consecutivos es menor.

Como resultado de la anarmonicidad, no solo es posible observar la banda **fundamental** ( $\Delta\nu = \pm 1$ ), sino también pueden observarse las bandas denominadas **sobretonos** (primer y segundo sobretono) correspondientes a transiciones  $\Delta\nu = \pm 2, \pm 3, \dots$

En el modelo anarmónico, los niveles de energía no están igualmente espaciados, a niveles superiores de energía el incremento de energía entre niveles consecutivos es menor, por lo que los sobretonos aparecen a frecuencias ligeramente menores que las correspondientes a múltiplos de las frecuencias fundamentales.

Además de los sobretonos, en la región NIR aparecen otro tipo de bandas de absorción. En las moléculas poliatómicas pueden darse cambios simultáneos en la energía de dos o más modos de vibración y la interacción entre ellos da lugar a las

bandas de **combinación** cuyas frecuencias son sumas o restas de múltiplos de cada una de las frecuencias que interaccionan:

$$\nu_{comb} = n_1\nu_1 + n_2\nu_2 + n_3\nu_3 + \dots \quad (1.6)$$

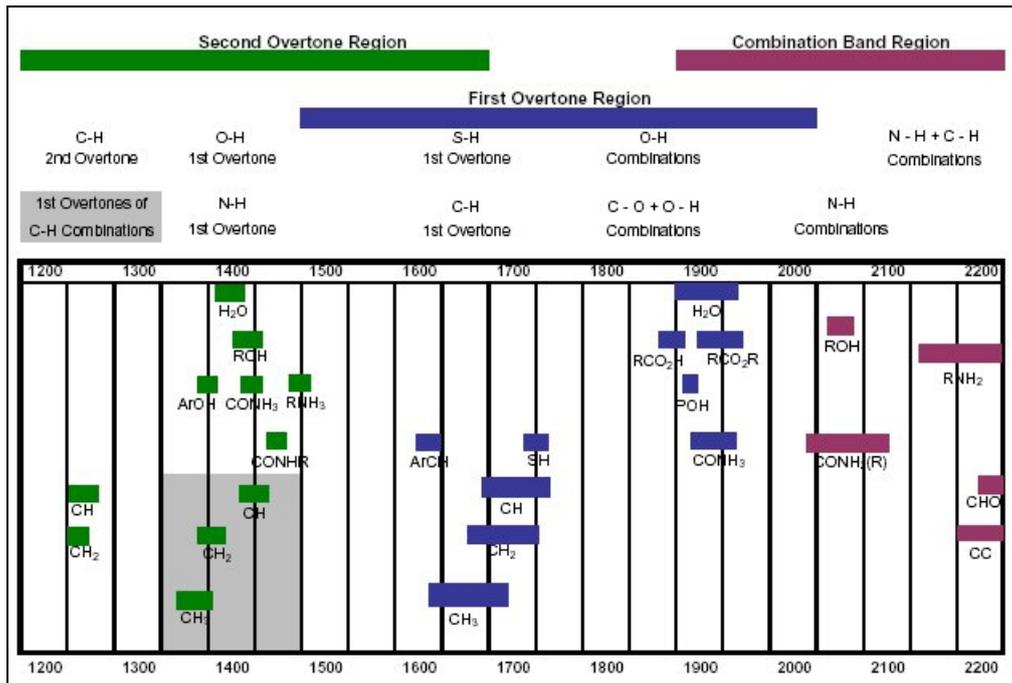
donde  $n_i$  son números enteros y  $\nu_i$  son las frecuencias de las transiciones que contribuyen a las bandas de combinación.

Las bandas de combinación más frecuentes son aquellas en que  $n_1=n_2=1$ .

El origen de la absorción de las bandas en la región NIR es el mismo que en la región IR; la molécula absorbe la radiación si la energía de ésta corresponde a la diferencia de energía entre dos niveles energéticos vibracionales de la molécula y si se produce un cambio en su momento dipolar.

Las bandas NIR no son tan bien definidas como en la región del IR medio, ya que son el resultado del solapamiento de sobretonos y bandas de combinación. Las bandas en esta región son poco intensas, ya que dependen de la anarmonicidad del enlace. El átomo de H es el más ligero, por lo que la vibración que contenga este átomo tendrá una mayor amplitud que en otro enlace donde hayan átomos más pesados. Al ser las vibraciones más amplias, las desviaciones del modelo del oscilador armónico serán mayores. Por esta razón, las principales bandas en la región del NIR son las debidas a los enlaces en los que intervienen átomos ligeros como C-H, O-H, N-H, S-H, ...

En la **Figura 4** se muestran las regiones del NIR donde absorben los diferentes enlaces, así como si pertenecen al primer, segundo o tercer sobretono o a combinaciones de frecuencias de vibración de diferentes enlaces.



**Fig. 4** Asignación de bandas en la región del infrarrojo cercano.

Los sobretonos del enlace C=O aparecen a longitudes de onda mayores que los sobretonos de los otros enlaces, debido a que la frecuencia fundamental de la vibración *stretching* del enlace C=O aparece a 1750-1550 cm<sup>-1</sup>, dependiendo del grupo funcional. La banda correspondiente al grupo -CHO que aparece a 2200 nm corresponde a la combinación de las vibraciones de los enlaces C-H y C=O del grupo aldehído.

### 1.3 Instrumentación

A finales de 1960, Karl Norris –considerado como el padre de la espectroscopia NIR–, demostró el valor potencial de esta región espectral en el análisis cuantitativo [5]. El

único instrumento comercial disponible era el desarrollado para medidas en las regiones del ultravioleta y visible, con la posibilidad de medir en el infrarrojo cercano como característica adicional. Norris y col. decidieron diseñar un instrumento para medir específicamente en la región del infrarrojo cercano.

Los instrumentos pueden clasificarse de diversas formas [5]. Los fabricantes suelen remarcar las ventajas de sus instrumentos, pero el consumidor tiene que ser lo más objetivo posible para hacer una correcta selección acorde con su problemática. Los instrumentos se diseñan o bien para un tipo específico de muestra o para hacer medidas diversas. En el trabajo de investigación la elección más apropiada, es un instrumento con amplias capacidades.

El esquema básico de un instrumento NIR es similar al de cualquier espectrofotómetro. Sus componentes son: fuente de radiación, sistema de selección de longitud de onda (SLO), compartimiento de muestra y detector.

### **1.3.1 Fuente de radiación**

La fuente de radiación más empleada es la lámpara de halógeno de filamento de tungsteno con ventana de cuarzo que proporciona un espectro continuo de 320-2500 nm. Otra fuente de radiación son los LED (*Light Emitting Diodes*) que emiten radiación en regiones estrechas del espectro [6]. Los instrumentos que incorporan este último sistema no requieren de un sistema de selección de longitudes de onda.

### 1.3.2 Sistema de selección de longitud de onda

Los instrumentos poseen un sistema para la selección de las longitudes de onda que es un factor importante para obtener un buen nivel de sensibilidad. En función del dispositivo empleado para la selección de longitudes de onda pueden clasificarse en: instrumentos de filtros, monocromadores e instrumentos de transformada de Fourier.

La selección de longitudes de onda mediante filtros se realiza interponiendo entre la fuente de radiación y la muestra, materiales que permiten el paso de determinadas longitudes de onda [5]. Los más sencillos son los de absorción que transmiten longitudes de onda de forma selectiva según el material de que están hechos. Otro tipo de filtros son los de interferencia óptica que filtran longitudes de onda según el índice de refracción del filtro y su grosor.

Los monocromadores son dispositivos que descomponen la luz policromática que proviene de la fuente de radiación en longitudes de ondas discretas. La radiación entra en forma de haz estrecho y un elemento dispersante (prisma o red de difracción) la descompone. Los más utilizados son las redes de difracción ya que son más baratos, proporcionan mejor separación de longitudes de onda y dispersan linealmente la radiación [7]. Otro monocromador es el llamado optoacústico (AOTF: *Acousto-Optic Tunable Filters*) que aprovechan el cambio de índice de refracción de un material ( $\text{TeO}_2$ ) cuando es atravesado por una onda sonora, convirtiéndose el material en una red capaz de difractar ciertas longitudes de onda de un haz incidente [8].

Los instrumentos de Transformada de Fourier dividen la luz en dos haces desfasados, recombinándolos posteriormente. De esta manera, los haces

recombinados pueden crear interferencias constructivas y destructivas. Mediante la Transformada de Fourier se convierte el interferograma (representación de la señal en dominio tiempo) en el espectro (en dominio frecuencias) [7].

### 1.3.3 Compartimiento de la muestra

Para registrar el espectro de una muestras, ésta se coloca en un compartimiento que se encuentra entre el sistema de selección de longitudes de onda y el detector. Dependiendo de la naturaleza de la muestra, ésta deberá ser presentada de distinta forma. Las muestras sólidas se miden en modo **reflectancia** y las muestras líquidas se miden por **transmisión** de la radiación. Un caso intermedio es la medida por **transflectancia**, en el que una parte de la radiación es reflejada y otra parte atraviesa la muestra y es reflejado por un material colocado en la cara opuesta de la muestra.

En la **Figura 5** se muestran los esquemas básicos de los diseños más comunes, que dependen del tipo de medida NIR (transmitancia, reflectancia, transflectancia).

Cuando la radiación interactúa con la muestra puede ser absorbida, transmitida o reflejada. En todos los casos, la señal NIR obtenida es una función compleja, de la que hay que extraer la información deseada. A continuación se exponen los fundamentos de cada tipo de medida NIR.

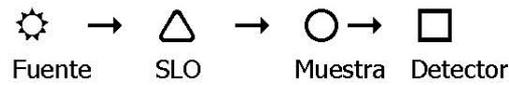
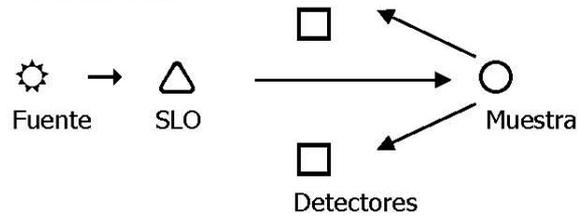
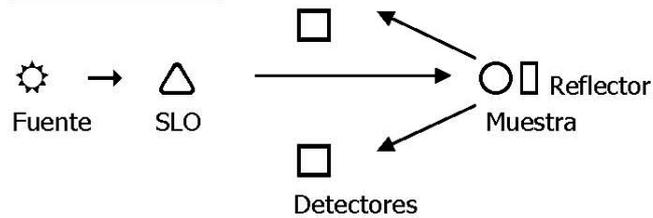
TransmisiónReflectanciaTransflectancia

Fig. 5 Diseños de los espectrómetros NIR.

**Medidas por transmisión**

La absorción de la radiación ( $A$ ) sigue la ley de Lambert-Beer [9]. La transmitancia de la solución ( $T$ ) es la fracción de radiación incidente transmitida por la solución. Ambas medidas se relacionan en la siguiente expresión:

$$A = \log \frac{1}{T} = \log \frac{P_0}{P} = abc \quad (1.7)$$

donde  $A$  es la absorbancia de la muestra,  $T$  la transmitancia,  $P_0$  la intensidad de la radiación incidente,  $P$  la intensidad de la radiación transmitida,  $a$  la absorptividad molar,  $b$  el camino óptico y  $c$  la concentración.

Pueden producirse desviaciones del cumplimiento de esta ley debidas a causas muy diversas (reflexión y/o dispersión de la radiación, concentraciones muy elevadas, interacciones químicas de la muestra, etc.)

### Medidas por reflectancia

La reflexión de la radiación puede ser especular o difusa. La reflectancia especular descrita por las leyes de Fresnel [10] no aporta información sobre la composición de la muestra, por lo que solo contribuye al ruido. La reflectancia difusa tiene lugar en todas las direcciones como consecuencia de los procesos de absorción y dispersión. Cuando los materiales son débilmente absorbentes a la longitud de onda incidente y cuando la penetración de la radiación es grande en relación con la longitud de onda, predomina la reflectancia difusa.

Debido a la complejidad del proceso de reflectancia difusa, no existen teorías rigurosas, sino aproximaciones entre las cuales la más importante es la de Kubelka-Munk [11].

Otra expresión empleada con mucha frecuencia es una aplicación empírica de la relación entre la concentración y la reflectancia relativa, similar a la ley de Lambert-Beer:

$$A = \log \frac{1}{R} = a'c \quad (1.8)$$

donde  $A$  es la absorbancia aparente,  $R$  la reflectancia relativa ( $R = R_{\text{muestra}}/R_{\text{estándar}}$ ),  $c$  la concentración y  $a'$  la constante de proporcionalidad.

### 1.3.4 Detectores

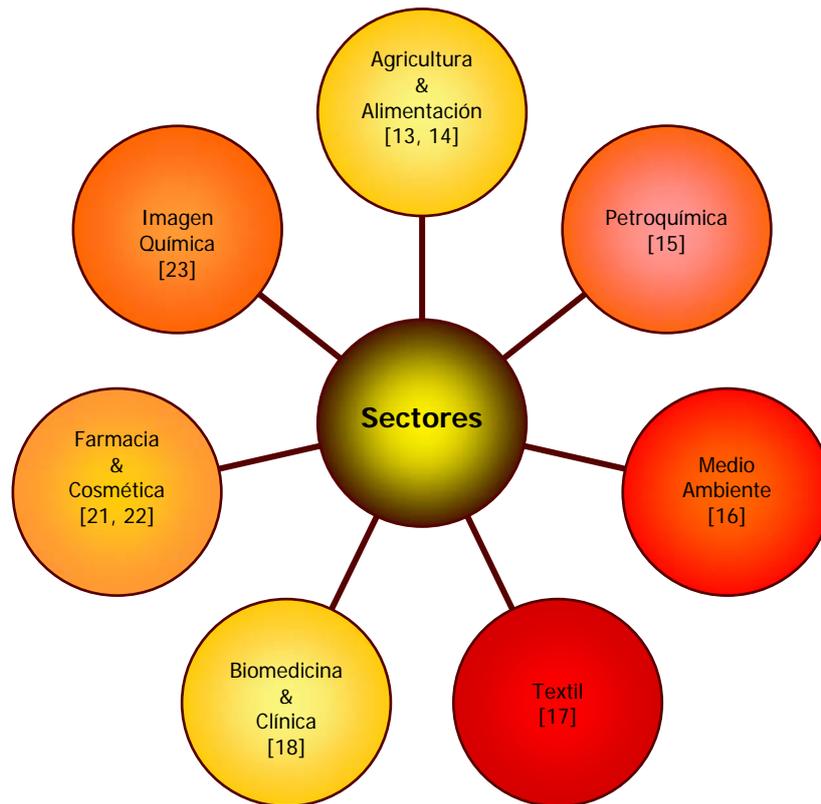
En la región del infrarrojo cercano, los detectores más habituales son los construidos con semiconductores (InGaAs, InAs, InSb, PbS, Si, ...). Hasta 1100 nm el semiconductor más empleado en los detectores es el Si y en la región entre 1100 y 2500 nm el material más utilizado es el PbS. Para medidas por transmisión en sólidos se utilizan detectores de InGaAs, que es operativo en la región de 600 a 1900 nm.

## 1.4 Aplicaciones de la espectroscopia en el infrarrojo cercano (NIRS)

La espectroscopia NIR se ha mostrado, en los últimos años, como una técnica analítica de gran interés en el control analítico de procesos industriales, en virtud de sus características tales como la no necesidad de reactivos, la escasa o nula preparación de la muestra, la rapidez de trabajo y la capacidad de proporcionar información de varios parámetros (analitos) simultáneamente.

El objetivo de este epígrafe es ilustrar el amplio campo de aplicación de la espectroscopia NIR en la industria, lo cual ha sido plasmado por algunos autores de forma satisfactoria [12].

En la **Figura 6** se muestran algunos ejemplos de la aplicación de la espectroscopia NIR en el análisis cualitativo y cuantitativo.



**Fig. 6** Aplicación de la espectroscopia NIR en diversos sectores.

---

El primer sector donde se empleó la espectrometría NIR fue en la agricultura. En 1974 la Comisión de Granos Canadiense adoptó esta técnica en sustitución del método tradicional (Kjeldahl) para determinar proteínas en trigo, lo cual supuso una disminución de los costes y redujo el vertido de productos [13].

D.C. Slaughter y col. [14] emplearon la espectroscopia NIR para la determinación no destructiva del contenido de sólidos solubles en frutas de papaya. El registro de los espectros NIR de la fruta se realizó con un espectrómetro NIR con sonda de fibra óptica y se construyó un modelo PLS para la determinación del contenido de sólidos solubles.

En las refinerías petroleras es cada vez más importante reducir los costes del proceso. Macho S. y Larrechi M.S discuten algunas consideraciones a tener en cuenta en la aplicación de la espectroscopia NIR en la determinación de propiedades de interés en la industria petroquímica [15]. Los autores, presentan dos ejemplos de aplicaciones: 1- determinación de hidrocarburos y densidad en muestras de nafta y 2- determinación de etileno y viscosidad en copolímeros polipropilénicos. Estas aplicaciones reflejan dos problemas diferentes: la nafta es la fuente de alimentación del proceso de craqueo y el análisis de nafta permite monitorear el proceso, sin embargo las propiedades físicas y químicas del polipropileno son analizadas en el control de calidad.

Un sector de gran auge en los últimos años es el del medio ambiente. El desarrollo de instrumentos portátiles permite llevar a cabo medidas in-situ. Stallard B.R. y col. desarrollaron una metodología mediante espectroscopia NIR para la determinación del grado de contaminación del suelo por aceites de motor [16].

La industria textil también se ha beneficiado de las ventajas de la espectroscopia NIR. El trabajo desarrollado por Cleve E. y col. presentan una variedad de aplicaciones de la espectroscopia NIR en la industria textil como son: la identificación de textiles, la determinación de humedad y el control del proceso [17].

El carácter no-invasivo de la espectroscopia NIR permite hacer diagnósticos sin necesidad de biopsias o cirugías. Rempel S.P. y col. presentan una reseña con 70 referencias bibliográficas, referentes al uso de la espectroscopia NIR y la imagen química en el campo médico, haciendo especial énfasis en los avances hechos en los

últimos años [18]. Se presentan aplicaciones ex-vivo e in-vivo que permiten ilustrar el potencial de esta técnica espectroscópica.

La fabricación de productos farmacéuticos es un sector industrial muy regulado tanto por organismos nacionales como por instituciones internacionales como son la European Medicines Agency (EMA) [19] y la International Conference on Harmonisation (ICH) [20]. La espectroscopia NIR es una técnica analítica que cumple muchas de las condiciones para su fácil y rápida aplicación en el control en la industria farmacéutica, ya que permite no sólo la caracterización de las diferentes formas de un principio activo sino también la determinación de la pureza polimórfica tanto del producto puro como de las preparaciones finales. Son cada vez más las compañías farmacéuticas que emplean la espectroscopia NIR como herramienta de análisis (cualitativo y cuantitativo) y alternativa a las técnicas habituales que generalmente son más laboriosas. Blanco M. y col. han publicado una revisión del uso de la espectroscopia NIR en la industria farmacéutica [21]. El trabajo incluye la discusión de un gran número de referencias relacionadas con el análisis cualitativo (identificación de materias primas y productos finales, estudios de homogeneidad, polimorfismo e isomería óptica) y cuantitativo (determinación de parámetros físicos, contenidos de humedad, principios activos y excipientes).

Cada año la industria cosmética invierte numerosos recursos para la creación y mejora de productos para la piel y el cabello, como son los limpiadores, hidratantes y acondicionadores. Muchas investigaciones están encaminadas a conocer la estructura y composición de la piel y el cabello, así como de las interacciones que se establecen entre estos y los productos elaborados. Kathleen M. presenta una descripción del

empleo de las técnicas espectroscópicas IR, Raman y NIR; en el estudio de propiedades estructurales y funcionales de piel y cabello [22]. En muchos casos estas técnicas pueden ser aplicadas directamente in-vivo o in-situ para obtener información del estado de la piel y el cabello, antes y después de aplicar los productos cosméticos.

La imagen química es un área del conocimiento relativamente nueva. La imagen NIR es la combinación de la espectroscopia NIR con el procesamiento de imagen digital. Un sistema de imagen NIR está compuesto por una fuente de iluminación, un sistema de imagen de óptica, un codificador espectral que selecciona las longitudes de onda y un plano focal de series lineales (*focal plane array*, FPA). El haz de luz proveniente de la fuente de iluminación del espectrómetro NIR incide sobre la muestra. La imagen de reflectancia difusa de la muestra es recogida por un sistema de imagen óptica, cuya configuración depende del tamaño y del tipo de muestra. Para las imágenes macroscópicas o microscópicas se utiliza una lente focal o un objetivo microscopio, respectivamente. La colección de datos se lleva a cabo registrando una serie de imágenes en el FPA del infrarrojo cercano (es decir InSb o InGaAs) en cada posición de la longitud de onda seleccionada mediante un codificador espectral. El resultado es un conjunto de datos tridimensionales, conocido como hipercubo espectral, donde el eje X e Y representan la información espacial y el eje Z representa la información espectral [23]. A pesar de las grandes ventajas que ofrece esta nueva herramienta de análisis, el número de trabajos que describen su uso práctico es limitado, siendo el campo de mayor aplicación la industria farmacéutica.

## Referencias

1. W. Herschel. *Philos. Trans. R. Soc.* **90**, 255-283 (1800).
2. W. Abney and E.R. Festing. *Philos. Trans. R. Soc.* **172**, 887-918 (1881).
3. W.W. Coblentz. *Investigations of Infrared Spectra*. Carnegie Institute Publication No. 35 (1905).
4. R.C. Nelson and W.R. Wilson. High resolving power infrared recording spectrometer. *Proceedings of the National Electronics Conference 3*, 579-86 (1947).
5. D.A. Burns and E.W. Ciurczak. *Handbook of Near-Infrared Analysis*, 2<sup>nd</sup> Ed by D.A. Burns and E.W. Ciurczak. Marcel Dekker, Inc. pp. 53-128 (2001).
6. A.S. Bonano and P.R. Griffiths. *Appl. Spectrosc.* **49**, 1590-1597 (1995).
7. D.A. Skoog, F.J. Holler, T.A. Nieman. *Principios de análisis instrumental*, 5<sup>a</sup> Ed by D.A. Skoog, F.J. Holler, T.A. Nieman. McGraw-Hill (2001).
8. C.D. Tran. *Anal. Chem.* **64**, 971A-981A (1992).
9. J.H. Lambert. *Photometria sive de mensura et gradibus luminis colorum et umbrae*, Augustae Vindelicorum (1760).
10. [http://en.wikipedia.org/wiki/Fresnel\\_equations](http://en.wikipedia.org/wiki/Fresnel_equations)
11. P. Kubelka and E. Munk. *Z. Tech. Phys.* **12**, 593-601 (1931).
12. M. Blanco and I. Villarroya. *TrAC, Trends Anal. Chem.* **21**, 240-250 (2002).
13. E. Stark. *Near Infrared Spectroscopy: The future Waves*, Ed by A.M.C. Davies and P. Williams. NIR Publications pp. 704 (1996).
14. D.C. Slaughter, C.G. Cavaletto, L.D. Gautz, R.E. Paul. *J. Near Infrared Spectrosc.* **7**, 223-228 (1999).
15. S. Macho and M. S. Larrechi. *TrAC, Trends Anal. Chem.* **21**, 799-806 (2002).

16. B.R. Stallard, M.J. Garcia, S. Kaushik. *Appl. Spectrosc.* **50**, 334-338 (1996).
17. E. Cleve, E. Bach, E. Schollmeyer. *Anal. Chim. Acta* **420**, 163-167 (2000).
18. S.P. Rempel and H.H. Mantsch. *Can. J. Anal. Sci. Spectros.* **44**, 171-179 (1999).
19. <http://www.emea.eu.int/pdfs/human/qwp/330901en.pdf>, Note for Guidance on the Use of Near Infrared Spectroscopy by the Pharmaceutical Industry (2003).
20. <http://www.ich.org/ich5q.html>, ICH Guidelines: Q2A Text on Validation of Analytical Procedures 1994, Q2B Validation of Analytical Procedures: Methodology (1996).
21. M. Blanco, J. Coello, H. Iturriaga, S. MasPOCH, C. de la Pezuela. *Analyst* **123**, 135R-150R (1998).
22. M. Kathleen. *Internet Journal of Vibrational Spectroscopy* [on-line computer file] **3**, No pp. given (1999).
23. G. Reich. *Adv. Drug Deliv. Rev.* **57**, 1109–1143 (2005).

# Capítulo 2

## 2. Quimiometría

### 2.1 Análisis Cualitativo

#### 2.1.1 Métodos de Reconocimiento de Pautas (PRMs)

##### 2.1.1.1 Métodos No Supervisados

##### 2.1.1.2 Métodos Supervisados

##### 2.1.1.3 Bibliotecas de Identificación

### 2.2 Análisis Cuantitativo

#### 2.2.1 Regresión Lineal Múltiple (MLR)

#### 2.2.2 Regresión en Componentes Principales (PCR)

#### 2.2.3 Regresión Parcial por Mínimos Cuadrados (PLS)

#### 2.2.4 Estrategias generales en el análisis

#### 2.2.5 Validación del método de análisis

## 2. Quimiometría

La Quimiometría es una disciplina que utiliza las matemáticas, estadísticas y lógica formal para: (1) diseñar o seleccionar los procedimientos experimentales óptimos y (2) proveer la máxima información química relevante a través del análisis de los datos [1].

La espectroscopia NIR es una técnica instrumental que proporciona un conjunto de datos que se define como señal analítica, la cual contiene información química y/o física de la muestra. Determinar la propiedad de interés analítico es un proceso complejo en la espectroscopia NIR, pues la señal es la contribución de más de una especie o propiedad. La aplicación de los métodos quimiométricos a la espectroscopia NIR proporciona una serie de ventajas ausentes en el análisis univariado.

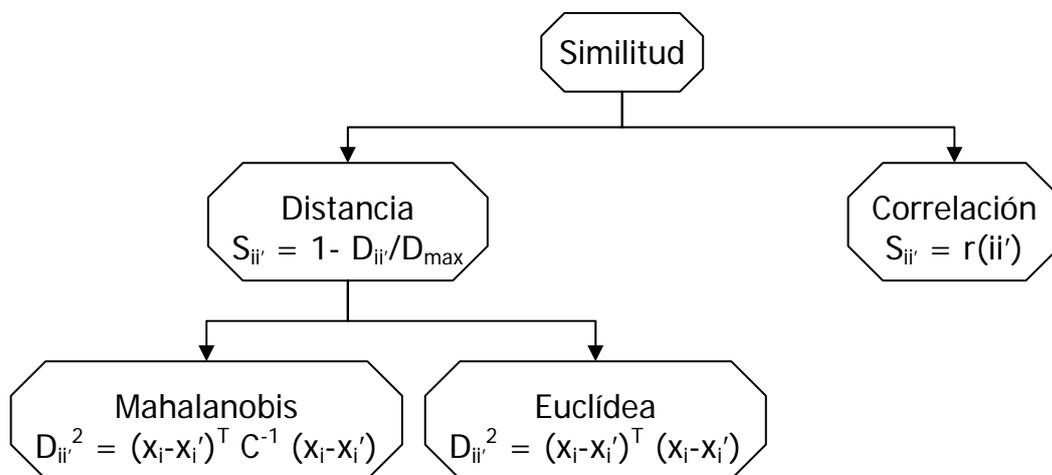
El objetivo de este epígrafe es resaltar los aspectos de mayor interés del análisis cualitativo y cuantitativo, empleando para ello diversas técnicas quimiométricas.

### 2.1 Análisis Cualitativo

Las técnicas quimiométricas empleadas en problemas de análisis cualitativo son conocidas como Métodos de Reconocimiento de Pautas (PRM's, *Pattern Recognition Methods*) [2].

### 2.1.1 Métodos de Reconocimiento de Pautas (PRM's)

Los PRM'S se aplican para clasificar objetos desconocidos en categorías o para separar objetos en categorías. La mayoría de estos métodos se basan en la medida de la similitud, que es la medida en que un objeto (espectro) es idéntico a otro. La similitud (S) entre objetos i e i', se expresa en términos de correlación o distancia, tal como se muestra en la **Figura 7**.

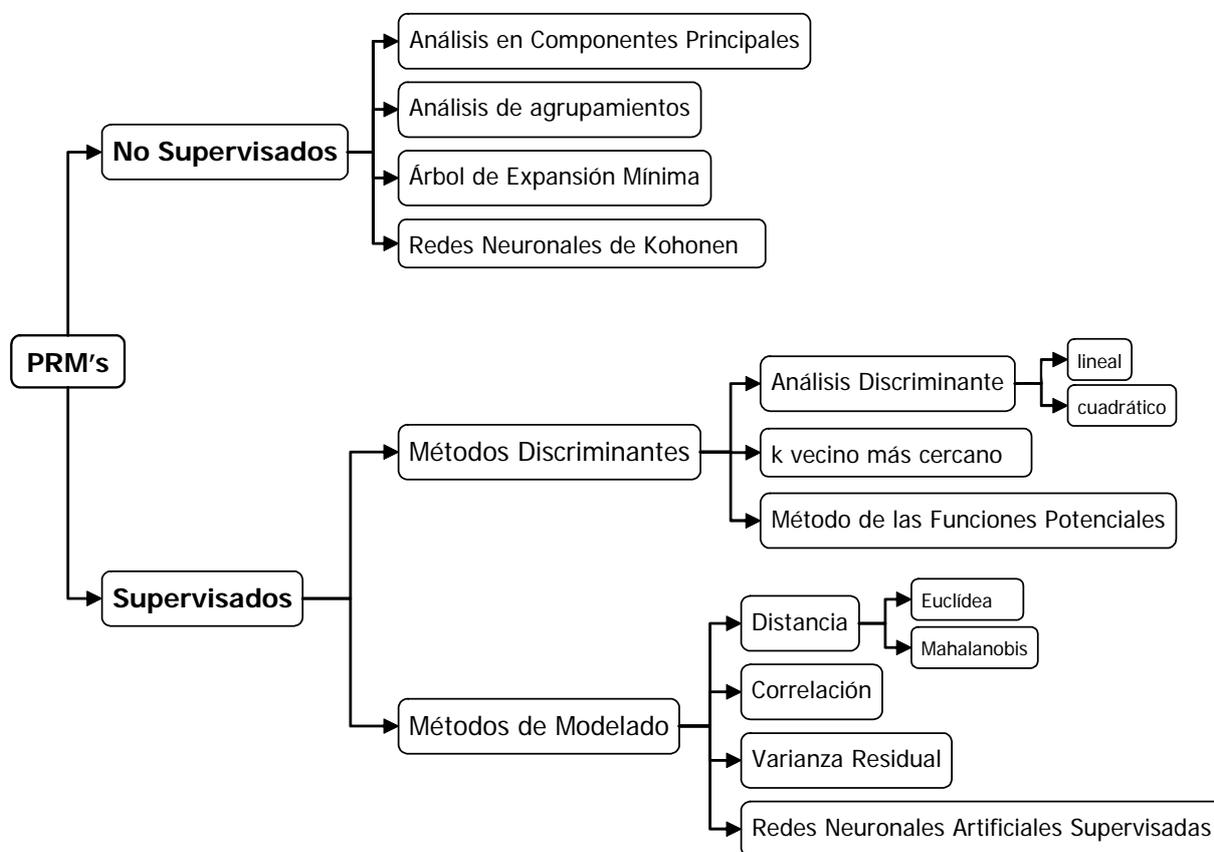


**Fig. 7** Criterios de similitud en los PRM's.

---

Los PRM'S se clasifican en métodos supervisados y no supervisados, dependiendo si se conoce o no, a priori, si el objeto pertenece a una clase determinada. Los métodos no supervisados buscan agrupamientos (espacio n-dimensional) sin conocer las clases a las cuales pertenece el objeto. En los métodos supervisados, se conocen las clases y se decide en cual de ellas un objeto debe ser clasificado. Es recomendable, incluso cuando el objetivo sea el desarrollo de un modelo supervisado, usar las técnicas no supervisadas para obtener un conocimiento

previo de los datos. En la **Figura 8** se resumen los métodos más relevantes los cuales se discutirán a continuación.



**Fig. 8** Clasificación de los PRM's.

Las técnicas quimiométricas utilizadas en problemás de análisis cualitativo en la presente tesis han sido el Análisis en Componentes Principales (PCA) para tener una información a priori de las características espectrales de las muestras y los métodos supervisados, en particular los de modelado para la clasificación de muestras.

### 2.1.1.1 Métodos No Supervisados

El objetivo de los métodos no supervisados es evaluar cuantos grupos existen en un conjunto de datos, sin información acerca de la pertenencia a una clase. Esto permite revelar para un conjunto de muestras los grupos naturales definidos por las medidas.

#### ☞ Análisis en Componentes Principales (PCA)

Los métodos instrumentales generan un gran volumen de información, por lo que uno de los aspectos más importantes dentro de la quimiometría es el desarrollo de métodos capaces de reducir este gran volumen de información a un número reducido de variables. Uno de los métodos de reducción de variables más empleados es el Análisis en Componentes Principales (PCA, *Principal Component Analysis*) que puede definirse como:

*El Análisis en Componentes Principales (PCA) es un método de compresión de datos que reduce un conjunto de datos registrados de  $M$ -variables y  $N$ -muestras a una representación más simple que emplea un menor número de "variables comprimidas" ( $A \ll M$ ), denominadas componentes principales (PCs) [3].*

El PCA busca las direcciones que explican la máxima variabilidad de las muestras, utilizándolas como nuevos ejes de coordenadas, denominados componentes principales (PC's). Los *loadings* son los cosenos de los ángulos que forman los nuevos ejes con los originales y los *scores* son las coordenadas de las muestras en los nuevos ejes.

La matriz  $\mathbf{X}$  (datos espectrales) se descompone en sus componentes principales,

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad (2.1)$$

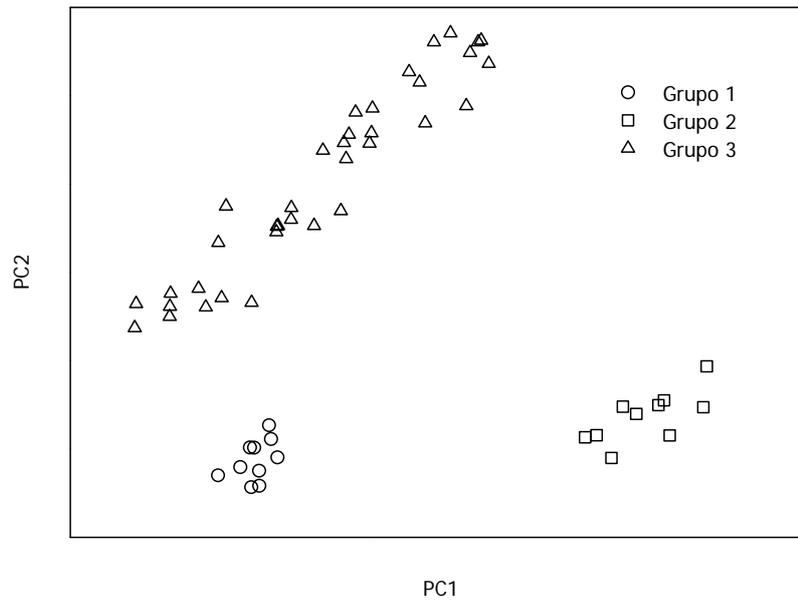
donde  $T$  es la matriz de *scores*,  $P$  la matriz de *loadings* y  $E$  la matriz de residuales.

Cada componente contiene información de diferente relevancia, así los primeros PC's describen la fuente de variación más importante de los datos. La representación completa de la matriz  $\mathbf{X}$  implica  $k$  vectores de *loadings* y *scores*, pero como al hacer la descomposición se reduce la dimensionalidad del sistema, la matriz  $\mathbf{X}$  resultante tiene  $a$  vectores donde  $a < k$ . La matriz  $\mathbf{X}$  resultante sería:

$$\mathbf{X} = t_1 p_1^t + t_2 p_2^t + t_3 p_3^t + \dots + t_a p_a^t + \mathbf{E} \quad (2.2)$$

No existe unanimidad en considerar el PCA como un método de clasificación, por el hecho de que no establece fronteras entre las diferentes clases ni aplica ningún criterio estadístico que permite discernir las clases. No obstante, es una técnica de reducción de variables que permite visualizar en un espacio de 2 ó 3 dimensiones, cuan similares o diferentes son un grupo de muestras entre si.

La interpretación de los resultados obtenidos con el PCA para una clasificación, se lleva acabo a partir de una representación del *score* de un PC frente al *score* de otro(s). 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. En la Figura 9 se muestra un gráfico de *scores* (PC2 vs. PC1) de un PCA de un conjunto de muestras.



**Fig. 9** Gráfico de *scores* (PC2 vs. PC1) de un PCA.

En la **Figura 9** se observa que las muestras se agrupan en tres grupos. Las muestras que conforman cada grupo son semejantes entre si; sin embargo, las muestras de los tres conjuntos se diferencian entre si.

### ☞ **Análisis de agrupamientos**

El análisis de agrupamientos se utiliza para clasificar objetos en grupos, los cuales están caracterizados por los valores de su conjunto de variables. Es una alternativa al PCA para describir la estructura de una tabla de datos.

Existen dos formas de representar los datos, la primera es en forma de un árbol, también llamado dendograma y consiste en la elaboración de una clasificación

jerárquica donde los grupos grandes se dividen otros más pequeños. Esta técnica examina la distancia entre los puntos de todas las muestras y representa esta información en un dendograma. La otra forma es la representación de los datos en una tabla que contiene diferentes agrupamientos. El agrupamiento es una partición en grupos y se denominan no jerárquicas [4].

### ☞ **Árbol de Expansión Mínima**

Esta técnica se basa en calcular la distancia entre un grupo de objeto. Los gráficos, que representan los objetos unidos mediante líneas de conexión que a su vez significan las distancias, se denominan árboles (*trees*). El gráfico para el cual la suma de las distancias es mínima se denomina *Minimal Spanning Tree* [5].

### ☞ **Redes Neuronales de Kohonen**

Las redes neuronales de Kohonen pertenecen a la clase de mapas auto-organizados y están diseñadas para tareas de clasificación.

Las redes Kohonen consisten en una capa de neuronas ordenadas en un mapa poco dimensional (organización uno dimensional o matriz bidimensional). Cada neurona o unidad contiene un vector peso de la misma dimensión que la estructura de entrada. Para entrenar la red, se aplica la regla de aprendizaje de Kohonen. Después del aprendizaje, cada uno de los vectores peso se orienta en tal dirección que la estructura de la topología se representa lo mejor posible en el mapa

resultante. Cada objeto o estructura de entrada se asigna a la neurona con el vector peso más similar. El objetivo de la red Kohonen es determinar los objetos que son similares o vecinos a una de las neuronas [6].

### **2.1.1.2 Métodos Supervisados**

Los métodos supervisados son usados cuando el objetivo es construir un modelo para clasificar futuras muestras como pertenecientes a una clase determinada.

Frecuentemente se hace una distinción entre los métodos basados en la discriminación y aquellos que se basan en el modelado de las clases. La mayoría de los métodos intentan encontrar de forma implícita o explícita, una frontera entre las clases.

### **Métodos Discriminantes**

Estos métodos establecen fronteras entre las clases y de esta forma clasifican un objeto como perteneciente a una clase o a otra, dependiendo del lado de la frontera en que se encuentra el objeto desconocido.

#### **☞ Análisis Discriminante**

Se basan en el concepto de que una función discriminante encuentra fronteras explícitas entre las clases y de esta forma el espacio queda dividido en regiones para

cada una de las clases. Los dos métodos más conocidos son el análisis discriminante lineal y el cuadrático [7].

#### ☞ **k vecino más cercano**

Es una técnica de clasificación que desde el punto de vista matemático es muy simple, que calcula la distancia entre el objeto desconocido y cada uno de los objetos pertenecientes a las clases. Se seleccionan los  $k$  objetos más cercanos al objeto desconocido y se aplica una regla de mayoría, donde el objeto queda clasificado en el grupo donde más vecinos tenga [8,9].

#### ☞ **Método de las Funciones Potenciales**

En este método cada objeto del conjunto de entrenamiento es considerado como un punto en el espacio rodeado por un campo de potencial. La clasificación de un objeto dentro de una clase se determina mediante el potencial acumulado de la clase en la posición del objeto desconocido. El potencial acumulado se calcula, sumando los potenciales individuales de los objetos de la clase en la posición del objeto desconocido. Finalmente el objeto se clasifica dentro de la clase que da el mayor potencial acumulado [10].

## Métodos de Modelado

Estos métodos determinan el espacio que ocupa cada clase, de forma que cada una de ellas será descrita por un modelo matemático. De esta forma un objeto desconocido puede clasificarse como pertenecientes a una de las clases o a ninguna de ellas.

### ☞ Distancia

El método PRIMA (*Pattern Recognition by Independent Multicategory Analysis*) se basa en el concepto de la **distancia Euclídea** [11]. Para cada clase se define una distancia que se basa en el centro de gravedad y no homogeneidad de la clase, Esta distancia es la que se emplea en la clasificación de objetos. Las condiciones para su aplicación son menos rigurosas que otros métodos y el algoritmo es simple.

El método UNEQ se aplica cuando solamente hay que considerar pocas variables [5]. Se basa en la **distancia de Mahalanobis** del centroide de la clase [12]. Cuando la distancia excede un valor de distancia, el objeto es considerado un *outlier* y por tanto no se considera que forma parte de la clase.

### ☞ Correlación

Los métodos basados en el coeficiente de correlación, calculan la correlación entre la muestra desconocida y las clases. Mientras mayor sea la correlación de la muestra

con la clase, mayor probabilidad de que esa muestra pertenezca a la clase y viceversa.

### ☞ Varianza Residual

Estos métodos se basan en el PCA de cada uno de los conjuntos, creando un modelo para cada uno de ellos [13]. Un espectro se reconstruye según cada uno de los modelos y los residuales que se obtienen son utilizados para calcular la probabilidad de que pertenezca a una clase o a ninguna de ellas. SIMCA (*Soft Independent Modelling of Class Analogies*) es el método más conocido de varianza residual [14].

### ☞ Redes Neuronales Artificiales Supervisadas

Describiremos las consideraciones básicas del funcionamiento de las redes neuronales en los procesos de clasificación supervisados [2].

Una red neuronal es un método de cálculo iterativo que permite ajustar los datos experimentales a una respuesta determinada mediante algoritmos de aprendizaje. Una de las redes más utilizadas son las *Multi-Layer Perceptron* (MLPs). La estructura de este tipo de redes consiste en una capa de entrada compuesta por tantos nodos como variables experimentales a modelar, la capa de salida consistente en la respuesta que se desee obtener y una capa oculta donde se realiza el proceso de aprendizaje. En el proceso de clasificación se han de establecer valores de salida indicativos de la pertenencia de un objeto a una clase o no.

### 2.1.1.3 Bibliotecas de identificación

Una de las aplicaciones de los PRM'S en la tesis es la creación de bibliotecas espectrales, en las cuales se emplean métodos supervisados de modelado. A partir de los espectros NIR de muestras de identidad conocida se crean distintas clases, lo cual permite identificar muestras desconocidas como pertenecientes o no a una de las clases predefinidas. La identificación se realiza comparando el espectro de la muestra desconocida con los espectros de las muestras que conforman la biblioteca.

Los métodos supervisados de modelado empleados en la construcción de las bibliotecas espectrales son:

- Correlación
  - Distancia Máxima
  - Distancia de Mahalanobis
  - Varianza Residual
- } basados en el espacio de las longitudes de onda
- } basados en el espacio de los componentes principales

En todos los métodos se busca un tipo y valor de umbral que define las fronteras de la clase. Para los métodos de distancia Mahalanobis y varianza residual, es posible calcular la probabilidad de un espectro dado, de pertenecer a la distribución representada por el espectro del producto. Por esta razón, para estos métodos existen dos tipos de umbral: valor de coincidencia (representa distancia o varianza residual) y nivel de probabilidad. Para los otros dos métodos (correlación y distancia máxima) solo es posible como tipo de umbral el valor de coincidencia.

Cuando se emplean los métodos de distancia Mahalanobis y varianza residual, se ha de definir el número de factores o varianza residual (%).

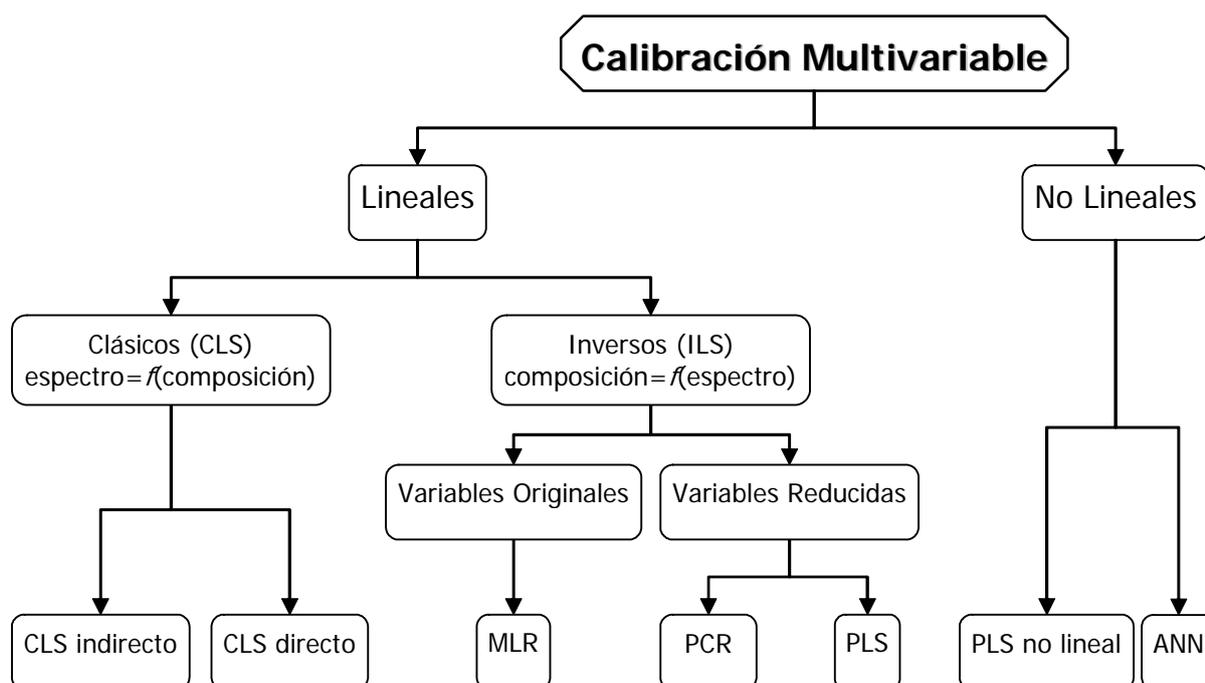
En este epígrafe hemos plasmado solamente los parámetros particulares de las bibliotecas espectrales, pues más adelante se discutirán otros parámetros a tener en cuenta en la construcción de las bibliotecas que son generales tanto para el análisis cualitativo como para el cuantitativo. De modo general, la cualificación es una etapa previa a la cuantificación.

## 2.2 Análisis Cuantitativo

En espectroscopia NIR, a diferencia de otras técnicas instrumentales, se obtienen más de una variable respuesta (matriz **X**) y estas a su vez pueden correlacionarse con propiedades físicas y químicas (matriz **Y**). Esto implica la necesidad de emplear métodos de calibración multivariantes, que permitan mediante un modelo matemático, correlacionar la matriz **X** con la matriz **Y**. Una vez establecido el modelo de calibración, se podrán predecir propiedades de interés analítico en muestras desconocidas a partir de sus correspondientes señales instrumentales.

Los métodos de Calibración Multivariantes pueden clasificarse siguiendo diferentes criterios [2], lo cual se representa de forma esquemática en la **Figura 10**.

La selección del método más apropiado dependerá de la problemática a analizar, del conocimiento previo del sistema y de los datos disponibles, entre otros. En la presente memoria solo discutiremos los métodos lineales, que afortunadamente toleran cierto grado de no linealidad.



**Fig. 10** Clasificación de los métodos de Calibración Multivariables.

Los métodos clásicos de mínimos cuadrados (CLS, *Classical Least Square*) asumen el cumplimiento de la ley de Lambert-Beer.

Los métodos inversos son los más empleados en la espectroscopia NIR por lo que a continuación discutiremos los aspectos más relevantes de cada uno de ellos. En ningún caso se harán las descripciones matemáticas de los métodos, por ser este un aspecto ampliamente detallado en la bibliografía [1-2, 15-16].

### 2.2.1 Regresión Lineal Múltiple (MLR)

El método inverso (ILS, *Inverse Least Square*) más simple es la Regresión Lineal Múltiple (MLR, *Multiple Linear Regression*) que asume una relación entre la

absorbancia y la concentración que es inversa a la ley de Lambert-Beer. En la ecuación resultante la concentración actúa como variable dependiente y la absorbancia a cada longitud de onda son las variables independientes.

La gran desventaja de este método es que el número de muestras debe ser mayor que el número de variables, por lo que solo se pueden emplear un número reducido de longitudes de onda. La selección de las variables es compleja y además la utilización de un número elevado de variables puede traer problemas de colinealidad, lo que disminuye la precisión de los resultados. Esto hace que este método sea poco utilizado en la espectroscopia NIR para el análisis cuantitativo.

### **2.2.2 Regresión en Componentes Principales (PCR)**

Los métodos inversos que emplean la reducción de variable (PCR, *Principal Component Regresión* y PLS, *Partial Least Square Regresión*), son los más empleados en la espectroscopia NIR y se basan en que la información contenida en las variables medidas se pueda concentrar en un número más pequeño de variables (componentes principales), sin perder la información relevante. El proceso de calibración se simplifica ya que el método PCR aprovecha las propiedades de la descomposición en componentes principales (PCA), realizando una regresión múltiple inversa (ILS) de la propiedad a determinar sobre los *scores* obtenidos en el PCA, en lugar de realizarla sobre los datos originales. Una de las ventajas de estos métodos es que se puede usar el espectro completo y no hay necesidad de seleccionar previamente un número determinado de variables respuesta.

Un caso particular es el PCR de pasos sucesivos (PCR *stepwise*), donde se añaden en pasos sucesivos los *scores* que más correlacionados están con la propiedad de interés analítico e incluso combinaciones de ellos [1]. Se construye una matriz  $\mathbf{X}$  con los *scores* ( $S_i$ ), el cuadrado de los *scores* ( $S_i^2$ ) y combinaciones de los *scores* ( $S_i S_j$ ), como se representa en la siguiente ecuación:

$$Y_{analitico} = b_0 + \sum_i^n b_i S_i + \sum_i^n b_{ii} S_i^2 + \sum_{i,j}^n b_{ij} S_i S_j \quad (2.3)$$

donde  $b_{ij}$  son los coeficientes de regresión y  $S_{ij}$  son los *scores* con  $i \neq j$ .

La inclusión de cada uno de los términos de la ecuación de regresión se evalúa mediante un análisis de varianza de la regresión (ANOVA). Para ello se calcula la F parcial según la expresión:

$$F_{parcial} = \frac{\left( \frac{SS^{con} - SS^{sin}}{df} \right)_{reg}}{\left( \frac{SS^{con}}{df} \right)_{res}} \quad (2.4)$$

donde  $SS$  es la suma de cuadrados y  $df$  son los grados de libertad.

La inclusión o no de un término se realiza según los siguientes criterios:

$F_{parcial} > F_{(0.05, df( Numerador ), df( Denominador ))}$  el término se incluye en el modelo.

$F_{parcial} < F_{(0.1, df( Numerador ), df( Denominador ))}$  el término se excluye del modelo.

### 2.2.3 Regresión Parcial por Mínimos Cuadrados (PLS)

En el método PLS a diferencia del PCR, intenta contener la mayor información para la predicción de las muestras en los primeros componentes principales. Para ello, durante la etapa de calibración, el algoritmo PLS utiliza tanto la información contenida en la matriz **X** como la información contenida en la matriz **Y**, obteniéndose unas variables auxiliares llamadas variables latentes, factores o componentes PLS.

Tanto en el PCR como en el PLS, la elección del número de componentes es un aspecto esencial durante el desarrollo de los modelos de calibración. Un número pequeño o elevado de factores en el modelo, implica modelos sencillos o complejos, respectivamente. Por esta razón es importante construir modelos que si bien deben ser lo más sencillo posible, deben tener un número suficiente de factores para que tengan una buena capacidad predictiva. Unos parámetros muy útiles para seleccionar el número de componentes son la varianza explicada (EV) y la varianza residual (RV), donde la primera aumenta y la segunda disminuye a medida que aumenta el número de factores. Si bien los criterios más empleados son el de Wold [17] y Haaland y Thomas [18], en la práctica se construyen modelos con variando el número de factores y se selecciona aquel que arroja una mejor capacidad predictiva.

### 2.2.4 Estrategias Generales en el Análisis

A continuación se describe la estrategia general que suele seguirse en la construcción de los modelos, que si bien es similar tanto para el análisis cualitativo como para el cuantitativo, cada uno de ellos tiene algunas particularidades.

1. **Seleccionar el modelo quimiométrico.** Teniendo en cuenta las características del sistema objeto de estudio y lo que se pretende determinar, se selecciona un modelo quimiométrico (métodos cualitativos y/o cuantitativos).
2. **Examinar los datos.** Hacer un análisis de los datos para ver si existen errores burdos. Estos errores pueden producirse tanto en las variables medidas (Ej. espectros) como en los valores de referencia (propiedades físicas y/o químicas), por lo que es importante examinar ambas variables.
3. **Selección de las muestras de los conjuntos de calibración y predicción.** Este es un aspecto de especial relevancia, ya que el conjunto de calibración debe estar constituido por muestras que cubran el rango del parámetro a determinar y también la variabilidad espectral. Una forma habitual de hacer esta selección es a partir del gráfico de los *scores* del PCA. También se ha de seleccionar un segundo conjunto de muestras que conformaran el conjunto de predicción, con las cuales se evaluará la capacidad predictiva del modelo construido.
4. **Pre-tratamiento.** Pueden existir fuentes de variación aleatorias o sistemáticas que enmáscaran la variación de interés. Estas variaciones indeseadas reducen la efectividad del modelo. La comprensión de la química y la física subyacente en este tipo de fuentes de variación, ayuda a la selección de un pre-tratamiento espectral adecuado. Es importante tener en cuenta que los pre-tratamientos modifican las variables originales y si no se aplican correctamente pueden eliminar variaciones de interés analítico. Dentro de los más empleados se pueden citar los siguientes:

- 
- **Promediado de espectros:** El ruido de alta frecuencia es aleatorio, por lo que el promediado de varias señales de una misma muestra reduce la contribución del ruido y por ende aumenta la relación señal/ruido.
  
  - **Suavizado espectral:** Se aplica cuando el promediado de espectros no es suficiente ya que la relación señal/ruido es pequeña. Los métodos más habituales son los basados en los filtros de Savitzky-Golay [19] y en transformadas de Fourier [20].
  
  - **Corrección de línea base:** Intenta corregir determinadas tendencias en la línea base que aumentan el ruido de la señal. Existen varios tipos de corrección de línea base, dependiendo del efecto que se desea corregir. En espectroscopia NIR el más empleado es el llamado ajuste de línea base (DT, *De-Trending*) que ajusta el espectro original a una función cuadrática [21].
  
  - **Derivadas:** Es uno de los pre-tratamientos más empleados en espectroscopia NIR pues reducen los problemas más frecuentes de esta técnica (solapamiento de bandas y variaciones de la línea base). La primera derivada elimina los términos constante a todas las longitudes de onda, es decir los desplazamientos de la línea base. La segunda derivada elimina los términos que varían linealmente con la longitud de onda. No suelen utilizarse derivadas de orden superior. Los métodos más empleados son los propuestos por Norris [22] y Savitzky-Golay [19].
  
  - **Corrección del efecto multiplicativo de la dispersión** (MSC, *Multiplicative Scatter Correction*): Corrige las diferencias espectrales debido al diferente tamaño de partículas (*scattering*) de muestras sólidas [23]. El

principal inconveniente de este método es que si el conjunto de espectros originales es modificado, el espectro de referencia cambia, por lo que los cálculos para aplicar el pre-tratamiento habría que hacerlos de nuevo.

- **Variable Normal Estándar** (SNV, *Standard Normal Variate*): Es otro pre-tratamiento que se aplica a la corrección del tamaño de partícula, pero con la ventaja que se aplica individualmente a cada espectro, sin depender de ningún espectro de referencia [21].

- **Corrección Ortogonal de la Señal** (OSC, *Orthogonal Signal Correction*): Es un método desarrollado por Wold *et. al.* [24] basado en una idea simple, que consiste en eliminar solamente la información espectral (matriz **X**) que no esta correlacionada (es ortogonal) con la matriz **Y**. El método se aplica a los espectros del conjunto de calibración y posteriormente se emplea esta corrección en la predicción de muestras nuevas.

5. **Rango espectral.** La selección del rango espectral no esta predefinida y depende de la situación particular. En el caso particular de la determinación de un analito, un posible rango espectral puede ser la región donde el analito tenga bandas significativas. Sin embargo, este criterio tiene el inconveniente de que puede eliminarse una región que contenga información importante de otras variables de las muestras. La elección es empírica y se hará en función de los resultados obtenidos.
6. **Evaluación de los Modelos.** La capacidad predictiva de un modelo se hace siempre con muestras externas que no han sido empleadas en el proceso de calibración.

En el análisis cualitativo la capacidad predictiva se evalúa a través del número de aciertos en la clasificación de muestras pertenecientes a una clase definida, así como el rechazo de muestras que no pertenecen a ninguna de las clases definidas.

En el análisis cuantitativo se compara el valor estimado (PRED) con el método de calibración desarrollado con el valor teórico o de referencia (REF). La diferencia entre estos dos valores es lo que habitualmente se conoce como residual. Existen numerosas maneras de expresar el error para un número "n" de muestras [16]. Dos de las expresiones más comunes para evaluar los errores de calibración y predicción son:

$$\text{(RSE, Relative Standard Error)} \quad RSE(\%) = \sqrt{\frac{\sum_{i=1}^n (PRED - REF)^2}{\sum_{i=1}^n (REF)^2}} \times 100 \quad (2.5)$$

$$\text{(RMSE, Root Mean Square Error)} \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (PRED - REF)^2}{n}} \quad (2.6)$$

### 2.2.5 Validación del método de análisis

Los métodos de análisis utilizados en el control de calidad, que no sean métodos oficiales de análisis, deben de ser debidamente validados *a priori* de su uso en rutina. La validación de un método de análisis tiene como objetivo, demostrar su idoneidad antes de su uso en rutina.

Existen numerosos textos que sirven de guía para la validación de los métodos analíticos. Una de las normativas más empleadas es la guía de la *International Conference on Harmonisation* (ICH) [25], sin embargo esta guía se redactó para la validación de métodos cromatográficos que eran los más empleados en el control de calidad. En los últimos años se ha producido un auge en la aplicación de la espectroscopia NIR [26] y una muestra de ello es que la *European Medicines Agency* (EMA) en el año 1991 elaboró unas guías para el uso y validación de procedimientos analíticos mediante Espectroscopia en el Infrarrojo Cercano en la industria farmacéutica [27].

En los trabajos que se presentan en esta memoria, las validaciones se han realizado siguiendo la normativa de las guías ICH y EMA evaluándose los siguientes parámetros: selectividad, linealidad, rango, exactitud, precisión (repetibilidad y precisión intermedia), robustez, límite de detección y cuantificación.

En la **Tabla 1** se muestra de forma esquemática los parámetros de calidad que suelen utilizarse en función del ensayo a realizar y a posteriori se detallan cada uno de ellos.

| <b>Tabla 1. Parámetros de Calidad.</b>   |                |                             |                   |        |
|--|----------------|-----------------------------|-------------------|--------|
| Parámetro  | Identificación | Determinación del contenido | Test de impurezas |        |
|  |                |                             | Cuantitativo      | Límite |
| Selectividad (1)   | +              | +                           | +                 | +      |
| Linealidad   | -              | +                           | +                 | -      |
| Rango  | -              | +                           | +                 | -      |
| Exactitud  | -              | +                           | +                 | -      |
| Precisión:   |                |                             |                   |        |
| ▪ Repetitividad  | -              | +                           | +                 | -      |
| ▪ Precisión Intermedia   | -              | + (2)                       | + (2)             | -      |
| Límite de Detección (LD)   | -              | -                           | - (3)             | +      |
| Límite de Cuantificación (LQ)  | -              | -                           | +                 | -      |
| Robustez   | +              | +                           | +                 | +      |
| + parámetro evaluado   |                |                             |                   |        |
| - parámetro NO evaluado  |                |                             |                   |        |
| (1) La falta de selectividad puede ser compensada con otros métodos de análisis. |                |                             |                   |        |
| (2) No es necesaria si se evalúa la reproducibilidad.                            |                |                             |                   |        |
| (3) Puede ser necesaria en algunos casos.  |                |                             |                   |        |

## ☞ Selectividad

La selectividad es la capacidad de un método para analizar el analito de interés y no otro componente. En espectroscopia NIR –a diferencia de otras técnicas instrumentales– este concepto se aplica de un modo diferente. Los modelos de calibración se desarrollan con muestras que contienen el analito a cuantificar y el resto de componentes de la muestra. Previa a la cuantificación se realiza un análisis cualitativo, que consiste en una identificación mediante una biblioteca espectral preparada con muestras que pertenecen a la misma clase. Con esto se garantiza que las muestras que van a ser cuantificadas, son las que han sido identificadas correctamente.

### ☞ **Linealidad**

La linealidad se demuestra obteniendo una respuesta proporcional entre la señal y la propiedad a determinar, en todo el intervalo. En las calibraciones univariadas la linealidad se obtiene una relación lineal entre la señal y la concentración. En las calibraciones multivariadas se evalúa la linealidad mediante la relación entre el valor de referencia y el valor estimado con el método desarrollado. Se recomienda tener un mínimo de 5 niveles de concentración que cubran todo el rango.

### ☞ **Intervalo**

Es el intervalo entre los niveles extremos de concentración, que dependerá de los estudios de linealidad, exactitud y precisión. Las guías ICH aconsejan cubrir intervalos de 80-120% para el análisis cuantitativo y 70-130% para los *tests* de uniformidad de contenido. En espectroscopia NIR el intervalo dependerá también en gran medida de la finalidad del método. En cada caso se debe evaluar que el intervalo establecido sea el adecuado.

### ☞ **Exactitud**

La exactitud se evalúa por la semejanza entre el valor estimado y el valor de referencia, ya sea este último obtenido por pesada o por un método de referencia. Las guías ICH recomiendan realizar un mínimo de 9 determinaciones cubriendo 3 niveles de concentración o bien 6 determinaciones al 100% de la concentración. La

exactitud quedara demostrada con un *test-t* de las diferencias, donde el promedio de estas diferencias deber ser significativamente igual a cero a un nivel de significación establecido.

### ☞ **Precisión**

La precisión es el error aleatorio o indeterminado asociado al método de análisis. Puede ser expresada como la desviación estándar relativa o coeficiente de variación (%). Existen tres niveles de evaluación:

- Repetitividad: Es la precisión de un método analítico bajo las mismas condiciones de operación en un espacio corto de tiempo. Se evalúa con un mínimo de 9 determinaciones a 3 niveles de concentración o con un mínimo de 6 determinaciones al 100% de la concentración.
- Precisión Intermedia: Expresa el grado de reproducibilidad de una serie de resultados, variando las condiciones normales de trabajo (días, operadores, temperatura, lotes de producción, etc.). Se evalúa la influencia de estos factores mediante un Análisis de Varianza (ANOVA) y se expresa el resultado con el coeficiente de variación global (%).
- Reproducibilidad: Generalmente implica estudios interlaboratorios, ampliando la precisión intermedia con factor adicional distintos laboratorios. Este parámetro de calidad es obligatorio para aquellos métodos que optan por el carácter de método oficial de la farmacopea.

### ☞ Límite de detección (LD)

Se define como la mínima concentración del analito que se puede detectar en una muestra, aunque no pueda ser cuantificada a ese nivel de concentración. Las guías ICH recogen las formas más comunes de estimar el LD.

### ☞ Límite de cuantificación (LQ)

Se define como la mínima concentración del analito que puede ser cuantificada con precisión y exactitud. Se suelen usar para su cálculo los mismos parámetros que para el límite de detección, pero con criterios ligeramente distintos.

### ☞ Robustez

La robustez es el parámetro que evalúa la influencia de pequeños cambios en las condiciones analíticas sobre la fiabilidad del método analítico. De una manera más amplia que la precisión intermedia, evalúa la susceptibilidad del método respecto pequeñas variaciones del procedimiento. En el caso de determinar que hay parámetros que afectan al método de análisis, estos deberán ser controlados y el procedimiento del método debe incluir recomendaciones para mantener bajo control estos parámetros.

## Referencias

1. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, S. Smeyers-Verbeke. *Handbook of Chemometrics and Qualimetrics: Part A*, Ed by B.G.M. Vandeginste and S.C. Rutan. Elsevier pp. 1-20 (1997).
2. B.G.M. Vandeginste, D.L. Mássart, L.M.C. Buydens, S. De Jong, P.J. Lewi and S. Smeyers-Verbeke. *Handbook of Chemometrics and Qualimetrics: Part B*, Ed by B.G.M. Vandeginste and S.C. Rutan. Elsevier pp. 207-241 (1998).
3. C.E. Miller. *Process Analytical Technology*, Ed by K.A. Bakeev. Blackwell Publishing Ltd pp. 226-328 (2005).
4. N. Bratchell. *Chemom. Intell. Lab. Syst.* **6**, 105-125 (1989).
5. O. Strouf. *Chemical Pattern Recognition*, Ed by O. Strouf. Research Studies Press Ltd (1986).
6. P. Cáceres-Alonso and A. Garcia-Tejedor. *J. Near Infrared Spectrosc.* **3**, 97-110 (1995).
7. P.A. Lachenbruch. *Discriminant Analysis*, Ed by P.A. Lachenbruch. Hafner Press (1975).
8. D. Coomans and D.L. Massart. *Anal. Chim. Acta* **136**, 15-27 (1982).
9. D. Coomans and D.L. Massart. *Anal. Chim. Acta* **138**, 153-165 (1982).
10. D. Coomans and I. Broeckaert. *Potential Pattern Recognition in Chemical and Medical Decision Making*, Ed by D. Coomans and I. Broeckaert. Research Studies Press Ltd (1986).
11. I. Juricskay and G.E. Veress. *Anal. Chim. Acta* **171**, 61-76 (1985).
12. M. P. Derde and D. L. Massart. *Anal. Chim. Acta* **184**, 33-51 (1986).
13. VISION, User Manual Foss NIRSystems (1998).

14. P.J. Gemperline, L.D. Webber, O. Frank. *Anal. Chem.* **61**, 138-144 (1989).
15. M. Blanco and I. Villarroya. *TrAC, Trends Anal. Chem.* **21**, 240-250 (2002).
16. K.R. Beebe, R.J. Pell, M.B. Seasholtz. *Chemometrics: A Practical Guide*, Ed by K.R. Beebe. John Wiley & Sons, Inc. pp. 245-339 (1998).
17. S. Wold. *Technometrics* **20**, 397-405 (1978).
18. D.M. Haaland and E.V. Thomas. *Anal. Chem.* **60**, 1193-1202 (1988).
19. A. Savitzky and M.J.E. Golay. *Anal. Chem.* **36**, 1627-1639 (1964).
20. G. Horlick. *Anal. Chem.* **44**, 943-947 (1972).
21. R.J. Barnes, M.S. Dhanoa, S.J. Lister. *Appl. Spectrosc.* **43**, 772-777 (1989).
22. K.H. Norris and P.C. Williams. *Cereal Chem.* **61**, 158-165 (1984).
23. P. Geladi, D. McDougall, H. Martens. *Appl. Spectrosc.* **39**, 491-500 (1985).
24. S. Wold, H. Antti, F. Lindgren, J. Ohman. *Chemom. Intell. Lab. Syst.* **44**, 175-185 (1998).
25. <http://www.ich.org/ich5q.html>, ICH Guidelines: Q2A Text on Validation of Analytical Procedures 1994, Q2B Validation of Analytical Procedures: Methodology (1996).
26. K.A. Bakeev. *Pharm. Technol. Europe* **15**, 27-28 (2003).
27. <http://www.emea.eu.int/pdfs/human/qwp/330901en.pdf>, Note for Guidance on the Use of Near Infrared Spectroscopy by the Pharmaceutical Industry (2003).

# Capítulo 3

## 3. Polimorfismo

3.1 Definición y propiedades

3.2 *Pseudo*-polimorfos. Formas amorfas, solvatos e hidratos

3.3 Métodos de caracterización

3.4 Efecto del procesamiento farmacéutico en fármacos polimórficos y solvatos

### 3. Polimorfismo

#### 3.1 Definición y propiedades

La capacidad de una sustancia de existir como dos o más formas cristalinas que tienen ordenación y/o conformaciones distintas de las moléculas en el enrejado cristalino se conoce como polimorfismo [1]. Esto –estrictamente– significa que los polimorfos son diferentes formas cristalinas de una misma sustancia pura.

Los sólidos polimórficos tienen diferente tamaño de celda y por ende diferentes propiedades físicas tales como: **empaquetamiento, termodinámicas, espectroscópicas, cinéticas, de superficie y mecánicas** [1]. Algunas de estas propiedades físicas tienen una relación directa sobre la efectividad terapéutica y estabilidad del medicamento, otras permiten la identificación y cuantificación de las distintas formas polimórficas.

Las diferencias en las dimensiones, configuración, simetría, capacidad (número de moléculas) y volumen vacío de las celdas unitarias de los diferentes polimorfos de una misma sustancia, hace que estos tengan diferentes propiedades físicas (volumen molar, densidad, índice de refracción, conductividad eléctrica y térmica, higroscopicidad) como consecuencia de las diferencias en el empaquetamiento molecular.

Cada polimorfo tiene su propia estructura cristalina y por ende su propia curva de energía potencial de Morse, donde existe una relación entre la energía de interacción molecular y la distancia intermolecular. La estabilidad relativa de los polimorfos y las fuerzas necesarias para la transformación polimórfica a temperatura

y presión constante, esta dada por la diferencia en la energía libre de Gibbs ( $\Delta G = \Delta H - T\Delta S$ ). La forma polimórfica más estable tiene la menor energía libre de Gibbs, fugacidad, presión de vapor, actividad termodinámica, solubilidad y velocidad de disolución por unidad de área en cualquier solvente, así como velocidad de reacción, incluyendo la velocidad de descomposición.

Las técnicas espectroscópicas basadas en las transiciones vibracionales, rotacionales y de spin nuclear, permiten la identificación y cuantificación de las formas polimórficas mediante diversas técnicas espectroscópicas (IR, Raman, RMN).

Las propiedades cinéticas tales como velocidad de disolución, velocidad de reacción en el estado sólido y estabilidad, son de gran importancia en la industria farmacéutica porque si bien es necesario que el medicamento este biodisponible, es necesario que permanezca inalterable en las condiciones ambientales.

Las propiedades mecánicas (dureza, fuerza de tensión, compactabilidad, *tableting*, *handling*, fluidez, mezclado) se deben a las diferencias en el empaquetamiento y en la energética de las interacciones moleculares.

### **3.2 Pseudo-polimorfos. Formas amorfas, solvatos e hidratos**

Muchos sólidos farmacéuticos presentan formas amorfas, que por sus características distintivas se tratan como polimorfos y a menudo se denominan *pseudo*-polimorfos. Las formas amorfas a diferencia de los verdaderos polimorfos, tienen una disposición desordenada de las moléculas y por lo tanto no poseen un enrejado cristalino distinguible, ni celdas unitarias [1]. Termodinámicamente, la ausencia de una energía

del enrejado estable provoca que la energía molar interna o la entalpía molar de las formas amorfas, excedan a las del estado cristalino. Además, la menor estabilidad y mayor reactividad de los sólidos amorfos, indica que la energía libre de Gibbs excede a la de las formas cristalinas.

Algunos sólidos farmacéuticos presentan diferente grado de solvatación con las moléculas del disolvente (agua, alcoholes, etc.) en el seno del cual se ha realizado el proceso de cristalización que son conocidos como solvatos y que no deben considerarse estrictamente como polimorfos y por ende también se les denomina *pseudo*-polimorfos.

Los cambios del estado de hidratación de las sustancias cristalinas (medicamentos y excipientes) pueden producirse durante todo el proceso de desarrollo. El comportamiento de hidratos farmacéuticos se ha convertido en objeto de atención durante la pasada década, sobre todo debido al impacto de éstos en el proceso de desarrollo y eficacia de la dosificación del medicamento. Las sustancias son hidratos/anhidros como resultado de cambios ambientales, en el procesamiento o por el paso del tiempo si es una forma termodinámicamente metaestable. Los hidratos cuando son sometidos a un proceso de deshidratación, se convierten en hidratos inferiores (menor número de moléculas de agua) e incluso anhidros.

### 3.3 Métodos de caracterización

Uno de los aspectos de mayor interés en el estudio de los polimorfos y *pseudo*-polimorfos, es la metodología analítica para llevar a cabo la caracterización de estos

compuestos. Byrn y col. definieron los tipos de formas sólidas que pueden encontrarse para un medicamento determinado [2].

**Polimorfos:** Formas que tienen la misma composición química, pero diferentes estructuras cristalinas.

**Solvatos:** Formas que contienen moléculas del solvente dentro de la estructura cristalina.

**Solvatos desolvatados:** Formas que cuando se les extrae el solvente de un solvato específico, mantienen la misma estructura cristalina.

**Amorfos:** Formas que no tienen una ordenación molecular de largo alcance.

De todos los métodos disponibles para la caracterización física de los compuestos sólidos, los más empleados se presentan en la **Tabla 2**.

| <b>Tabla 2. Métodos para la caracterización de los polimorfos.</b> |   |
|--|---|
| Métodos  | Técnicas Instrumentales                   |
| Estructurales  | Difracción de rayos X de simple haz       |
|  | Difracción de polvo de rayos X (XRPD)     |
| Morfológicos   | Microscopia de luz polarizada             |
|  | Microscopia térmica                       |
| Térmicos   | Termogravimetría (TG)                     |
|  | Análisis térmico diferencial (DTA)        |
|  | Calorimetría diferencial de barrido (DSC) |
| Espectroscópicos   | Infrarrojo (IR)                           |
|  | Raman                                     |
|  | Resonancia Magnética Nuclear (NMR)        |

Los métodos que se presentan en la **Tabla 2** permiten la caracterización de sólidos y por ende asegurar la existencia de una forma polimórfica.

La difracción de rayos X tiene un importante inconveniente para su utilización en rutina debido a su lentitud, necesidad de preparación de la muestra, etc. La industria necesita disponer de técnicas rápidas que puedan aplicarse al control del proceso y de los productos acabados.

La espectroscopia NIR es una técnica instrumental que cumple muchas de las condiciones para su fácil y rápida aplicación en este tipo de controles, ya que permite no sólo la caracterización de las diferentes formas de un principio activo sino también la determinación de la pureza polimórfica tanto del producto puro como de las preparaciones finales.

### **3.4 Efecto del procesamiento farmacéutico en fármacos polimórficos y solvatos**

La conversión de los polimorfos y la desolvatación de los hidratos, son fenómenos indeseados que pueden ocurrir durante el procesamiento farmacéutico. Condiciones ambientales de elevada humedad y temperatura, no son inusuales durante la fabricación en rutina de los medicamentos. Frecuentemente, las diferencias entre las temperaturas de fusión de los distintos polimorfos cristalinos son inferiores de 1kcal/mol, con una temperatura de transición por debajo de la temperatura de ebullición del agua. En el caso de los hidratos, remover el agua de la estructura cristalina requiere más energía, pero depende mucho más de las condiciones de temperatura y humedad a que ha sido sometida la sustancia. A continuación discutiremos como son posibles una gran variedad de conversiones de fases cuando

se exponen los sólidos a condiciones energéticas como son el almacenaje a granel, secado, molido, granulación húmeda, secado en estufa y compactación.

#### ☞ **Producción y almacenaje a granel**

El primer momento donde puede ocurrir un cambio en la forma polimórfica o grado de solvatación, es en el paso final de la cristalización en la síntesis del medicamento a granel. Existen numerosos parámetros que pueden afectar el proceso de cristalización como son: composición del solvente y polaridad, concentración del medicamento y grado de sobresaturación, temperatura y velocidad de enfriamiento, presencia de cristales madre y/o sitios de nucleación, aditivos que alteran el hábito del cristal o añaden tensión al enrejado cristalino, agitación, pH o la presencia de moléculas formadoras de sales. Existe la posibilidad de que ocurran cambios en la forma del polimorfo cuando el proceso es escalado, movido a un sitio nuevo o ejecutado por un nuevo operador.

#### ☞ **Reducción del tamaño de partícula**

El último paso en la producción del medicamento a granel, implica un molido para reducir la distribución del tamaño de partícula. Esto se lleva a cabo usando las condiciones más delicadas posibles para obtener muestras homogéneas o empleando molidos más rigurosos con el objetivo de reducir el tamaño de partículas primarias y de esta forma mejorar la homogeneidad de la formulación, disolución y biodisponibilidad. En el último paso del proceso se emplea una cantidad mayor de

energía, lo cual puede traer como resultado cambios polimórficos o la generación de sustancias amorfas. Siempre que se empleen procesos de molidos es necesario un análisis cuidadoso para detectar los posibles cambios en la cristalinidad o forma del cristal.

### ☞ Granulación

Cuando una formulación requiere de un paso de granulación, este representa una ocasión para que se produzcan cambios en el cristal del medicamento. El problema se acentúa cuando es una granulación húmeda, debido a que la adición a la mezcla del principio activo y excipientes de un solvente, potencia la transformación polimórfica. Las mezclas se homogenizan en fase húmeda y posteriormente se secan a una temperatura determinada. La granulación mejora los procesos de mezclado y fluido, permitiendo la compresión de tabletas a altas velocidades.

La granulación húmeda puede ser considerada como una suspensión del medicamento en una mezcla de solventes y excipientes. Si tenemos en cuenta que un solvente habitual es el agua, se pueden obtener una serie de conversiones entre anhidros e hidratos o entre hidratos con diferentes números de moléculas de agua. Es evidente que la transformación será siempre hacia la forma más estable.

### ☞ Secado

Cuando se seca una formulación que contiene el medicamento, existe la posibilidad que se produzcan cambios en la estructura cristalina. Las formas anhidras y amorfas se forman por simple desolvatación de un solvato y estas reacciones pueden ocurrir durante la etapa de secado.

### ☞ Compactación

Frecuentemente se asume que la forma cristalina de un medicamento se mantiene inalterable durante la etapa de compactación. El fundamento de esta hipótesis es que los excipientes blandos se deforman preferiblemente antes que las formas cristalinas que son más duras. Al aumentar las velocidades de *tableting* disminuye el tiempo de exposición a altas presiones disminuye. En la producción de la mayoría de los medicamentos esto es cierto, sin embargo existen numerosos ejemplos que no siguen esta regla general y se producen cambios en la forma del polimorfo o del solvato durante la etapa de compactación.

Como se puede apreciar, existen múltiples etapas durante la fabricación donde se pueden producir transformaciones polimórficas. Es importante que durante la etapa de desarrollo de un fármaco, los investigadores estudien minuciosamente la posibilidad del fármaco de presentar formas polimórficas. La elección de la forma termodinámicamente más estable es lo más aconsejable, pero en caso de no ser esta la forma de mayor biodisponibilidad, se necesita disponer de herramientas de análisis

para tener un control riguroso de la estabilidad y poder detectar posibles transformaciones durante el proceso de producción y almacenaje.

**Referencias**

1. *Polymorphism in Pharmaceutical Solids*, Ed by H.G. Britain. Marcel Dekker Inc. (1999).
2. S.R. Byrn, R.R. Pfeifer, M. Ganey, C. Hoiberg, G. Poochikian. *Pharm. Res.* **12**, 945- (1995).

## METODOLOGÍA Y DISCUSIÓN DE RESULTADOS

# Capítulos 4 & 5

4. Introducción

5. Metodología

5.1 Descripción y preparación de las muestras

5.2 Instrumentación

5.3 *Software*

5.4 Procesamiento de datos

#### 4. Introducción

La espectroscopia en el infrarrojo cercano es una técnica rápida y fiable que permite reducir el tiempo de análisis y asegurar la calidad del producto. Por otra parte, el espectro NIR contiene no solamente información química que se emplea para la determinación de la composición, sino también información física que permite determinar características físicas de las muestras. Sin embargo, esta técnica tiene algunas desventajas como son la baja selectividad, sensibilidad y una señal compleja que requiere el uso de técnicas multivariantes para extraer la información de interés analítico.

El objetivo de esta tesis es el estudio de la influencia de parámetros físicos en el desarrollo de modelos de calibración mediante la espectroscopia en el infrarrojo cercano y la influencia de los parámetros físicos sobre los espectros NIR, puede complicar la construcción de los modelos de calibración. Algunas de estas situaciones se estudian en la presente tesis.

La memoria consta de dos partes, una primera parte donde el efecto de la estructura cristalina de los compuestos (polimorfismo) sobre los espectros NIR hace que los polimorfos se comporten espectralmente como compuestos distintos, lo que se aplica a la determinación de mezclas de formas polimórficas de un fármaco (azitromicina) y una segunda parte en que se estudian diferentes estrategias para la corrección del efecto de la temperatura sobre los espectros NIR, usando como modelo mezclas de los compuestos implicados en un proceso de esterificación. En la primera parte, la diferencia en la estructura de los polimorfos, permite la identificación y cuantificación de uno en presencia de otro. Sin embargo, en la

segunda parte una propiedad física como es la temperatura tiene un efecto no deseado sobre el espectro NIR que afectan negativamente la capacidad predictiva de los modelos para la determinación del contenido de los componentes de las muestras.

En los siguientes epígrafes se indican los detalles referentes a la preparación de las muestras, los instrumentos empleados así como el registro de los espectros NIR, los programas utilizados y finalmente el procesamiento de los datos.

## 5. Metodología

### 5.1 Descripción y preparación de las muestras

- ❖ Las formas polimórficas de la azitromicina utilizadas en el desarrollo de los modelos son: amorfa (A), monohidrato (MH) y dihidrato (DH). Se prepararon en el laboratorio mezclas binarias de A+DH y MH+DH por pesada en balanza analítica en el rango de 0-10% de DH y se homogenizaron manualmente. Para comprobar la homogeneidad de las muestras se registraron los espectros NIR de la misma y cuando los espectros consecutivos eran idénticos, se consideraba la muestra homogénea. Los valores de las pesadas fueron tomados como valores de referencia. Con el objeto de observar el efecto de la temperatura y la presencia de contenidos variables de DH y humedad, se mantuvieron las muestras en diferentes condiciones: 1) ambiente de laboratorio, 2) en desecador con  $P_2O_{5(s)}$  y 3) en estufa a  $60^{\circ}C$  con 100% de humedad.

- ❖ Para la determinación de los compuestos implicados en el proceso de esterificación, se prepararon en el laboratorio por pesada en balanza analítica 9 mezclas binarias, 4 ternarias y 14 cuaternarias, de ácido acético glacial (HAc), 1-butanol (BuOH), n-butilo acetato (BuAc) y agua, en un rango de fracción molar (C) comprendido entre 0-0.8. Las mezclas (binarias, ternarias y cuaternarias) y los compuestos puros se sometieron a un calentamiento gradual a partir de 25 °C hasta 90 °C y se registraron los espectros NIR a intervalos de 5 °C. El recipiente con las disoluciones se coloca sobre una placa calefactora, se le introduce un agitador magnético y se sumerge un termómetro y la sonda de fibra óptica de inmersión para registrar los espectros NIR.

## 5.2 Instrumentación

Los registros de los espectros NIR de las muestras se han realizado con dos módulos de registro, que fueron seleccionados en función del tipo de muestra.

El registro de los espectros de las formas polimórficas de la azitromicina se ha realizado en un Espectrofotómetro FOSS NIRSystems, Modelo 6500 provisto de un módulo Rapid-Content Analyzer (RCA), en un intervalo de longitudes de onda de 1100-2500 nm. El registro de estos espectros se lleva a cabo en modo **reflectancia**; las muestras se colocan en una cubeta de vidrio, se remueven con una espátula y se registran (resultado de 32 scans) por triplicado en el rango de 1100-2500nm con una resolución de 2 nm. El espectro de referencia se obtiene con la cubeta de vidrio vacía y una placa cerámica incorporada al instrumento. La media de los tres espectros se utiliza para los posteriores tratamientos matemáticos.

El registro de los espectros de las muestras líquidas fueron realizadas en modo **transflectancia** en un espectrómetro FOSS NIRSystems 5000 equipado con una sonda de fibra óptica. El camino óptico se ajusta para que la absorbancia medida esté siempre dentro de la zona lineal de respuesta (inferior 1.5 ua). Los espectros de todas las muestras se registraron (resultado de 32 scans) en el intervalo de longitudes de onda comprendido entre 1100 y 2500 nm con una resolución de 2 nm. El espectro de referencia se mide con la sonda de fibra óptica en el aire antes de empezar el calentamiento de cada disolución.

En ambos espectrofotómetros NIR se realiza la comprobación del correcto funcionamiento del instrumento mediante los *test* de Linearización de longitud de onda y de Funcionamiento. Estos *test* se hacen diariamente antes de comenzar el registro de los espectros de las muestras.

La caracterización de las formas polimórficas de la azitromicina se realiza por difracción de rayos X, utilizando un Difractómetro Seifert XDR 3000 T/T, trabajando con: tubo de Cu ( $\lambda = 1.5418 \text{ \AA}$ ) a 50 kV y 40 mA, goniómetro MZVI con controlador de micropasos C3000, monocromador secundario de grafito y detector de centelleo geometría Bragg-Brentano. Los difractogramas se registraron en el rango de  $2\theta = 5^\circ - 25^\circ$ , con una amplitud de paso de  $0.02^\circ$  y un tiempo de irradiación de 10 segundos en cada posición.

La determinación de agua en las muestras de azitromicina se realizó mediante un valorador Karl-Fischer Methrom mod. 716 DMS Titrino. Se minimiza el tiempo transcurrido entre la valoración y el registro del espectro NIR de cada muestra, para evitar la posible absorción de humedad ambiental.

### 5.3 Software

Diferentes programas fueron empleados en esta tesis, dependiendo del objetivo que se pretendía. A continuación se presentan los programas y los algoritmos que en cada uno de ellos se utilizó.

- ❖ VISION v2.51 (FOSS NIRSystems, Silver Spring, MD, USA): Permite el control del instrumento, obtención y gestión de los espectros NIR, así como la construcción de las bibliotecas de identificación, que permiten llevar a cabo el análisis cualitativo.
- ❖ UNSCRAMBLER v9.1 (Camo Process S.A., Trondheim, Norway). Incorpora algoritmos que permiten aplicar los pre-tratamientos de primera y segunda derivada de Savitzky-Golay, el análisis en componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS).
- ❖ SPSS v11.5 (SPSS Inc., Chicago, IL, USA). Se empleó para la construcción de los modelos de calibración usando la regresión MLR *stepwise*.
- ❖ SIMCA-P v8.0 (Umetrics AB, Umeå, Sweden). Incorpora los algoritmos de los pre-tratamientos espectrales Standard Normal Variate (SNV) y Orthogonal Signal Correction (OSC), el análisis en componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS).

#### 5.4 Procesamiento de datos

Las características de la señal en espectroscopia NIR, hace necesario el uso de métodos multivariantes para extraer la información cualitativa o cuantitativa. El primer paso es elegir el método teniendo en cuenta el objetivo del análisis y una vez elegido el método multivariante se siguen una serie de pasos hasta la obtención de un modelo con una correcta capacidad predictiva (ver epígrafe 2.2.4).

En esta tesis se han empleado dos algoritmos de regresión, el de regresión parcial por mínimos cuadrados (PLS) y el de regresión lineal múltiple en pasos sucesivos (MLR *stepwise*). La regresión PLS es un algoritmo de uso frecuente en espectroscopia NIR, sin embargo, la regresión MLR *stepwise* no lo es. La relación no lineal entre los *scores* del PCA de las muestras registradas a diferentes temperaturas, indujo a pensar que una regresión que incluyera términos cuadráticos, podría modelar el efecto de la temperatura y por tanto se ensayaron modelos MLR *stepwise* con términos cuadráticos de los *scores* y con productos de los *scores*.

En cada regresión existen unos parámetros propios que deben ser ajustados además del pre-tratamiento espectral y el rango de longitudes de onda.

Los modelos PLS se construyeron por cross-validation (*leave-one-out*) seleccionando el número de factores correspondiente al mínimo en el gráfico de varianza explicada *vs.* el número de factores PLS.

Los modelos basados en el algoritmo MLR *stepwise*, se construyeron con los cinco primeros *scores* ( $S_i$ ), los cuadrados de los *scores* ( $S_i^2$ ) y sus productos binarios ( $S_i S_j$ ):

$$Y_{analito} = b_0 + \sum_i^5 b_i S_i + \sum_i^5 b_{ii} S_i^2 + \sum_{i,j}^5 b_{ij} S_i S_j + \quad (5.1)$$

donde,  $b_{i,j}$  son los coeficientes de la regresión y  $S_{i,j}$  son los *scores*.

La inclusión de cada uno de los términos de la ecuación de regresión se evalúa mediante un análisis de varianza de la regresión (ANOVA). Para ello se calcula la F parcial según la expresión:

$$F_{parcial} = \frac{\left( \frac{SS^{con} - SS^{sin}}{df} \right)_{reg}}{\left( \frac{SS^{con}}{df} \right)_{res}} \quad (5.2)$$

donde, SS es la suma de cuadrados y df son los grados de libertad.

La inclusión o no de un término se realiza según los siguientes criterios:

$F_{parcial} > F_{(0.05, df( Numerador), df( Denominador))}$  el término se incluye en el modelo.

$F_{parcial} < F_{(0.1, df( Numerador), df( Denominador))}$  el término se excluye del modelo.

Las características del espectro NIR hacen necesario de modo general, la aplicación de pre-tratamientos espectrales que reducen las contribuciones indeseadas de la señal. En el caso ideal las variaciones de la señal están relacionadas únicamente con la propiedad de interés analítico. Desafortunadamente esto prácticamente no ocurre y generalmente la señal es el resultado de la composición química (la concentración de todos los componentes) y de las propiedades físicas (tamaño de partículas, presión, temperatura, etc.) de la muestra. Habitualmente a los espectros (matriz **X**) se les aplican pre-tratamientos espectrales (SNV, derivadas,

OSC, etc.). Las derivadas de primer y segundo orden se obtuvieron usando el algoritmo de Savitzky-Golay con un polinomio de orden 2 y una ventana de 11 puntos. En el algoritmo OSC es necesario seleccionar el número de factores-OSC (generalmente con 1 ó 2 es suficiente) que eliminen o reduzcan en mayor medida la variabilidad espectral no correlacionada con la propiedad de interés analítico.

La selección del rango espectral en los modelos de calibración es empírica. Si se desea determinar la concentración de un analito, un posible rango espectral es aquel donde el analito tenga las bandas más significativas. El rango espectral que arroje los menores errores de calibración y predicción, será el utilizado para la construcción de los modelos de calibración.

Otro aspecto de gran importancia es la detección de muestras anómalas (*outliers*) que son muestras que presentan desviaciones en los modelos respecto al resto de las muestras y que tienen un efecto perjudicial sobre los modelos de calibración. Los gráficos de *scores* de un PCA es una herramienta útil para la detección de muestras que se encuentran muy alejadas del resto. Los gráficos de residuales e influencia (*leverage*) de la regresión PLS son también útiles en la detección de muestras con desviaciones. En la práctica cuando se ha identificado a una muestra como un *outlier*, se intenta siempre que sea factible, repetir el registro del espectro NIR y/o la determinación del valor de referencia.

Una vez encontrado un modelo con parámetros de calibración aceptables, se comprueba la capacidad predictiva del mismo con muestras externas que no han sido incluidas en la calibración y de las cuales se conoce la identidad (análisis cualitativo) o el valor de referencia (análisis cuantitativo). Los resultados de la

calibración y la predicción se evalúan mediante herramientas estadísticas que permiten comparar los resultados obtenidos con la espectroscopia NIR con los de referencia.

Existen unas guías (ver Epígrafe 2.2.5) para la validación de la metodología desarrollada y su aplicación al análisis en rutina en la industria.



# Capítulo 6

## 6. Discusión de los Resultados

### 6.1 Caracterización y análisis de formas polimórficas

#### 6.1.1 Introducción

#### 6.1.2 Caracterización de *pseudo*-polimorfos por difracción de Rayos X

#### 6.1.3 Análisis de *pseudo*-polimorfos mediante espectroscopia en el infrarrojo cercano

##### 6.1.3.1 Modelos de calibración (PLS)

##### 6.1.3.2 Estudio de Estabilidad

### 6.2 Estudio y corrección del efecto de la temperatura en espectroscopia en el infrarrojo cercano

#### 6.2.1 Introducción

#### 6.2.2 Estudio del efecto de la temperatura sobre los espectros NIR

#### 6.2.3 Estrategias para la corrección del efecto de la temperatura

##### 6.2.3.1 Modelar el efecto de la temperatura

##### 6.2.3.2 Eliminar el efecto de la temperatura

## 6. Discusión de los Resultados

Este capítulo de la tesis tiene como objetivo recoger los aspectos más relevantes de las metodologías de análisis desarrolladas. Se pretende dar una visión más detallada de cada uno de las metodologías y hacer una discusión de los avances que ellas representan.

### 6.1 Caracterización y análisis de formas polimórficas

En esta memoria se presenta un estudio de diferentes formas polimórficas de un fármaco (azitromicina). Se desarrollan modelos de calibración que permiten determinar una forma polimórfica en presencia de otra. Complementariamente, se lleva a cabo un estudio de estabilidad de las formas polimórficas.

#### 6.1.1 Introducción

La caracterización y determinación de las formas polimórficas de un fármaco tiene un gran interés en la industria farmacéutica. Los métodos más frecuentemente usados (Ej. la difracción de rayos X) son laboriosos y lentos y por consiguiente poco apropiados para su aplicación al análisis en la industria.

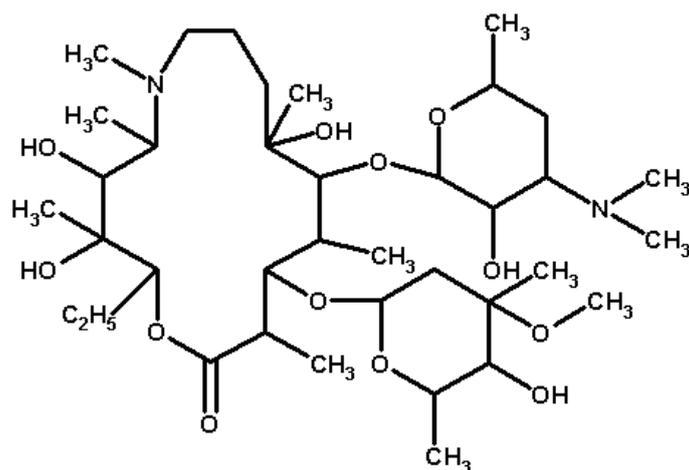
Si bien es importante poder caracterizar los diferentes polimorfos o *pseudo*-polimorfos de una sustancia, también es necesario poder detectar posibles transformaciones que pueden ocurrir en el proceso de producción, envasado y/o

preservación. Las diferencias en la estructura de los polimorfos conduce a diferencias en propiedades de interés farmacéutico, como es la biodisponibilidad, estabilidad química, etc.

La azitromicina es el primero de una familia de antibióticos macrólidos derivados de la eritromicina A. Se han descrito dos formas cristalinas que corresponden a un mono- y dihidrato (MH y DH) y también una forma amorfa anhidra (A), que si bien son tratados como polimorfos se denominan *pseudo*-polimorfos, ya que no cumplen estrictamente la definición de polimorfismo (ver epígrafe 3.1). El peso molecular de cada una de las formas *pseudo*-polimórficas son diferentes ( $PM_A=749$ ,  $PM_{MH}=767$ ,  $PM_{DH}=785$ ), por lo que desde un punto de vista riguroso no podemos considerar las tres formas como una misma sustancia. Por otra parte, la forma amorfa tiene una disposición desordenada de las moléculas y por lo tanto no poseen un enrejado cristalino distinguible.

En la **Figura 11** se representa la fórmula molecular de la azitromicina. Las fórmulas moleculares de las formas hidratadas (MH y DH) son similares con la única diferencia que tienen una y dos moléculas de agua de hidratación, respectivamente.

Las formas cristalinas (MH y DH) poseen un contenido teórico de agua de 2.3% y 4.6%, respectivamente; que corresponden al agua de hidratación. Las formas A y MH son ligeramente higroscópicas y su contenido de agua puede variar entre 1-3% sin que se alteren la forma *pseudo*-polimórfica.

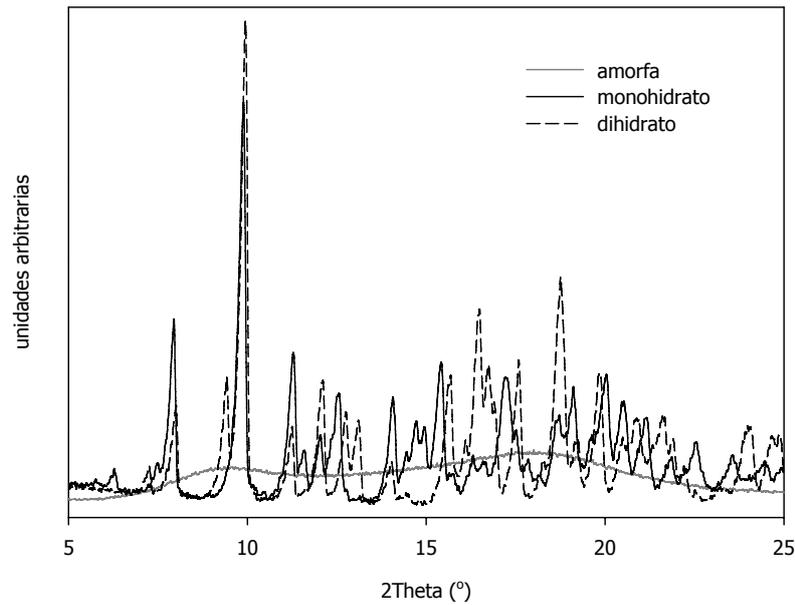


**Fig. 11** Estructura molecular de la azitromicina ( $C_{38}H_{72}N_2O_{12}$ ).

### 6.1.2 Caracterización de *pseudo*-polimorfos por difracción de Rayos X

Las tres formas *pseudo*-polimórficas fueron caracterizadas por difracción de Rayos X (Figura 12). Se puede apreciar como las formas cristalinas (MH y DH) presentan perfiles de difracción bien definidos con bandas agudas, mientras que la forma amorfa presenta bandas amplias y suaves debidas al *scattering* de los rayos X de la muestra.

Se calcularon mediante calibración univariada los límites de detección y cuantificación de la determinación de DH en muestras de A y MH. El procedimiento seguido se detalla para la determinación de DH en muestras de MH.



**Fig. 12** Difractogramas de las tres formas *pseudo*-polimórficas de la azitromicina.

Se registraron los difractogramas de mezclas binarias de MH+DH con contenidos de DH entre 0 y 5%. Se calculó el valor de  $2\theta$  para el cual la correlación entre la intensidad y el contenido de DH (%) era máxima. El mayor coeficiente de correlación (0.99) se obtuvo para  $2\theta=13.055^\circ$ . En la **Figura 13** se muestra el gráfico de regresión a  $2\theta=13.055^\circ$ , así como los parámetros de la recta.

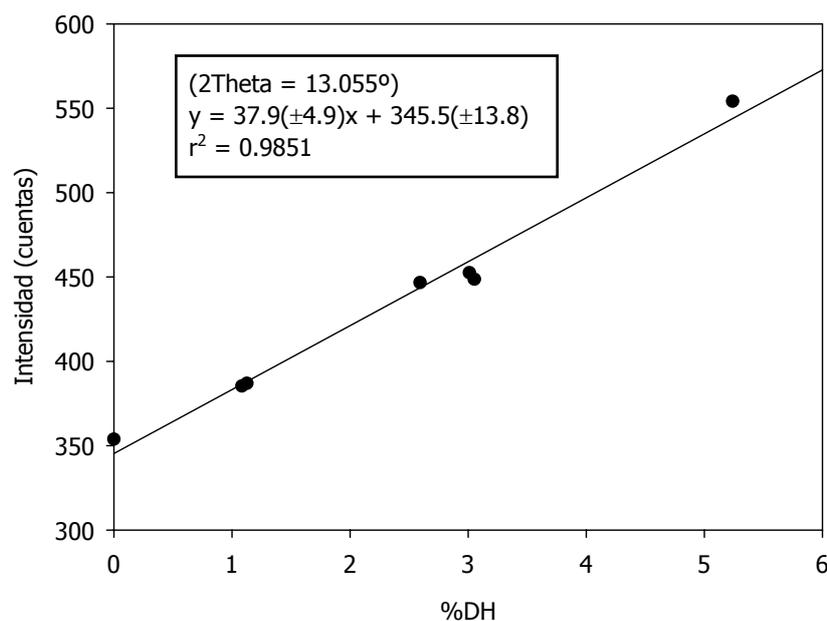
Los límites de detección y cuantificación se calcularon mediante las siguientes expresiones:

$$LD = \frac{3 \cdot SD_a}{b} \quad (6.1)$$

$$LD = \frac{10 \cdot SD_a}{b} \quad (6.2)$$

donde  $SD_a$  es la desviación estándar del intercepto y  $b$  es el valor de la pendiente.

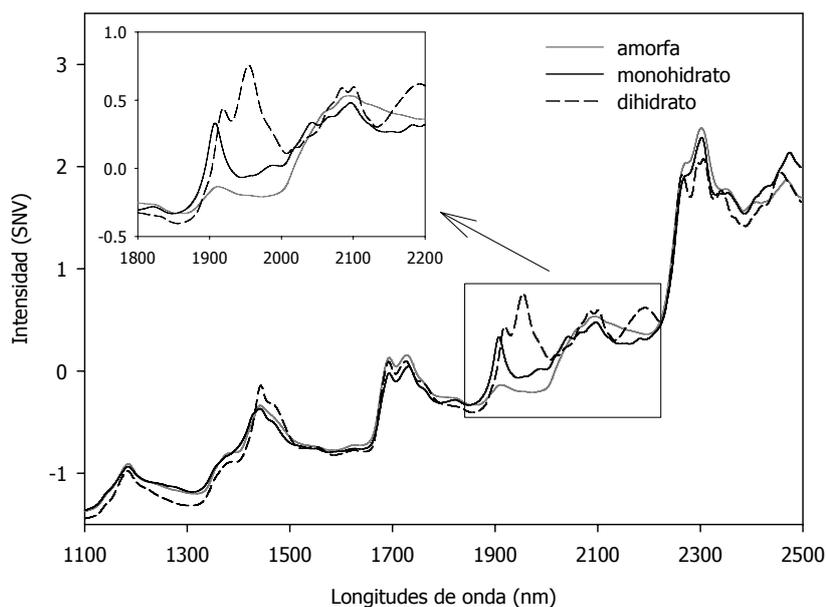
Los valores obtenidos de límites de detección y cuantificación fueron 0.5% y 1.5% respectivamente y serán comparados más adelante con los que se obtienen mediante la metodología desarrollada con la espectroscopia NIR.



**Fig. 13** Gráfico de regresión de Intensidad (cuentas) vs. %DH para  $2\theta=13.055^\circ$ .

### 6.1.3 Análisis de los *pseudo*-polimorfos mediante espectroscopia en el infrarrojo cercano

Las diferencias en la estructura de las tres formas *pseudo*-polimórficas, también se ponen de manifiesto en las diferencias de sus espectros NIR (Figura 14).



**Fig. 14** Espectro NIR de las tres formas *pseudo*-polimórficas de la azitromicina.

Los espectros NIR de las tres formas *pseudo*-polimórficas muestran las principales diferencias en la región de 1800-2200 nm donde se encuentra la banda más intensa del agua (ver **Figura 14**). La forma A tiene una banda ancha y poco intensa que corresponde al agua de adsorción, sin embargo las formas MH y DH tienen bandas intensas con uno y dos máximos de absorbancia, respectivamente que corresponden al agua de hidratación.

### 6.1.3.1 Modelos de calibración (PLS)

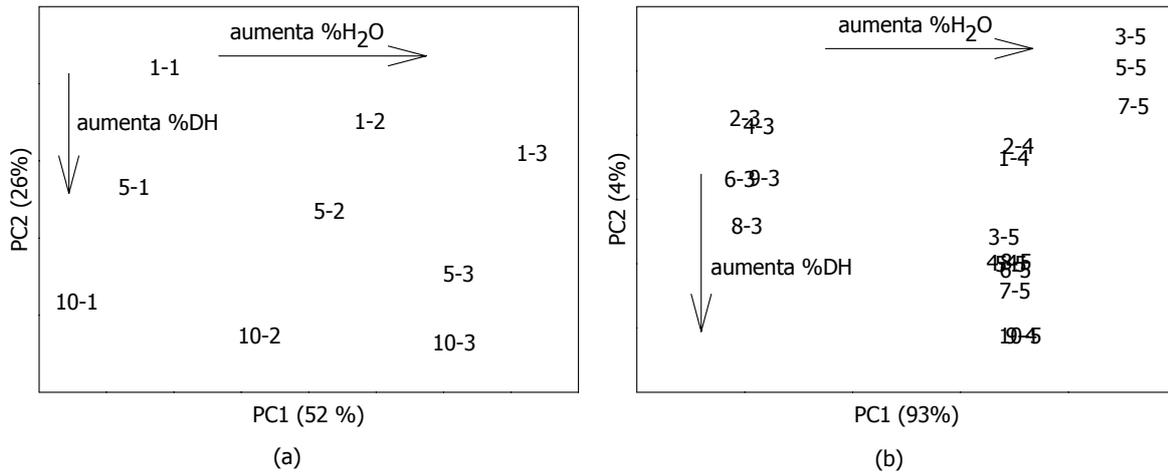
Las diferencias espectrales de las formas *pseudo*-polimórficas que se manifiestan en la **Figura 14**, permiten desarrollar modelos de calibración mediante regresión parcial

por mínimos cuadrados (PLS) para la determinación del contenido de DH en muestras de A y MH. También se desarrollaron modelos PLS para la determinación del contenido de agua en dichas muestras.

La posible absorción de humedad en los *pseudo*-polimorfos, conduce a la necesidad de estudiar este efecto y su posible influencia en los modelos de calibración para la determinación del contenido de DH.

Se prepararon mezclas de A y MH con contenidos variables de DH y se mantuvieron en diferentes condiciones de humedad: 1- estufa de vacío a 50°C, 2- desecador y 3-atmósfera húmeda. De esta forma se disponía de muestras con contenidos variables de DH y también de humedad. En la **Figura 15** se representan los gráficos de *scores* de un análisis en componentes principales (PCA) de estas muestras.

Los gráficos del *scores* del análisis en componentes principales (PCA) de la **Figura 15**, muestran que en ambas mezclas (A+DH y MH+DH) la mayor variabilidad queda explicada en el primer PC y está relacionada con el contenido de humedad. La variabilidad que explica el segundo PC, está relacionada con la variación del contenido de DH. Estos hechos indican que para la construcción de modelos de calibración, es necesario incluir en los conjuntos de calibración mezclas polimórficas con contenidos variables de humedad. El rango de concentración de DH establecido para la construcción de los modelos de calibración fue de 0-10%.



**Fig. 15** Gráficos de *scores* (PC1 vs. PC2) del PCA de mezclas de: (a) A+DH, (b) MH+DH. El primer número de las muestras representa el contenido de dihidrato (%DH) y el segundo el contenido de humedad (%H<sub>2</sub>O).

Los valores de referencia del contenido de DH se obtienen directamente por la pesada en la balanza analítica y los del contenido de agua por valoración con el método de Karl-Fischer.

Siguiendo las estrategias generales descritas en el epígrafe 2.2.4, se seleccionaron las muestras que conforman los conjuntos de calibración y de predicción, se ensayaron diferentes intervalos de longitudes de onda, pre-tratamientos espectrales, número de factores PLS y se evaluaron los resultados obtenidos.

En la **Tabla 3** se presentan los parámetros de calibración y predicción de los modelos PLS1 para las determinaciones de DH y agua, en los *pseudo*-polimorfos A y MH. La alta correlación entre la intensidad de las bandas de absorción de DH y H<sub>2</sub>O

en el rango 1800-2200 nm, nos llevó a ensayar la aplicación de modelos PLS2 para la determinación conjunta de estos dos analitos, siendo los resultados muy similares a los obtenidos con los modelos PLS1.

**Tabla 3.** Figura de méritos de los modelos PLS1 desarrollados para la determinación de dihidrato y agua en muestras de amorfa y monohidrato.

|                             | DIHIDRATO                        | AGUA                              | DIHIDRATO                        | AGUA                              |
|-----------------------------|----------------------------------|-----------------------------------|----------------------------------|-----------------------------------|
| Muestras                    | amorfa                           | amorfa                            | monohidrato                      | monohidrato                       |
| Rango de concentración (%)  | 0 – 10.1                         | 1.2 – 4.8                         | 0 – 13.3                         | 2.9 – 5.4                         |
| Rango espectral (nm)        | 1800–2200                        | 1800–2200                         | 1100–2500                        | 1800–2200                         |
| Pretratamiento espectral    | 2ª Derivada                      | 1ª Derivada                       | 2ª Derivada                      | 2ª Derivada                       |
| No. de Factores PLS         | 2                                | 2                                 | 4                                | 2                                 |
| Calibración                 |                                  |                                   |                                  |                                   |
| No. de muestras             | 30                               | 17                                | 39                               | 33                                |
| Recta de regresión          | $C_{NIR} = 0.073 + 0.983C_{REF}$ | $C_{NIR} = 0.017 + 0.994C_{REF}$  | $C_{NIR} = 0.069 + 0.986C_{REF}$ | $C_{NIR} = 0.097 + 0.977C_{REF}$  |
| Coefficiente de correlación | 0.992                            | 0.997                             | 0.993                            | 0.988                             |
| RSEC (%)                    | 7.8                              | 3.1                               | 6.8                              | 2.5                               |
| Predicción                  |                                  |                                   |                                  |                                   |
| No. de muestras             | 42                               | 25                                | 44                               | 26                                |
| Recta de regresión          | $C_{NIR} = 0.048 + 1.008C_{REF}$ | $C_{NIR} = -0.032 + 1.017C_{REF}$ | $C_{NIR} = 0.058 + 0.988C_{REF}$ | $C_{NIR} = -0.124 + 1.029C_{REF}$ |
| Coefficiente de correlación | 0.992                            | 0.993                             | 0.989                            | 0.977                             |
| RSEP (%)                    | 7.9                              | 5.2                               | 7.8                              | 2.7                               |

De los resultados de la **Tabla 3** cabe destacar que en la mayoría de los casos con el rango espectral de 1800-2200nm se obtienen los mejores modelos, que como se mostró en la **Figura 12** es donde se encuentra la banda más intensa del agua y donde existen las mayores diferencias espectrales entre las tres formas *pseudo*-polimórficas.

El modelo que determina el contenido de DH en muestras de MH, arroja los mejores resultados empleando el rango espectral de 1100-2500nm y además

necesita un mayor número de factores-PLS que el resto de los modelos. Esto puede deberse a la mayor similitud espectral que existe entre las dos formas cristalinas.

Se validaron ambas metodologías siguiendo las guías de validación descritas en el epígrafe 2.2.5 para su aplicación en la industria farmacéutica (Tabla 4).

**Tabla 4.** Resultados de las validaciones de las metodologías desarrolladas para la cuantificación del contenido de dihidrato en muestras de amorfa y monohidrato.

|                               |                                    | AMORFA  | MONOHIDRATO   |
|-------------------------------|------------------------------------|---|---|
| Selectividad                  | Método                             | Varianza Residual en el espacio de los PC                                   | Varianza Residual en el espacio de los PC                                   |
|                               | n                                  | 36  | 43  |
|                               | Rango de concentración (%)         | 0-2   | 0-2   |
|                               | Pretratamiento espectral           | 1ª Derivada   | 2ª Derivada   |
|                               | Rango espectral (nm)               | 1800-2200   | 1100-2500   |
|                               | Umbral                             | 0.90  | 0.74  |
| Linealidad                    | n                                  | 9   | 8   |
|                               | Rango de concentración (%)         | 0-10  | 0-10  |
|                               | Intercepto                         | -0.09 ± 0.19  | -0.19 ± 0.44  |
|                               | Pendiente                          | 1.04 ± 0.04   | 1.03 ± 0.08   |
|                               | Coefficiente de correlación        | 0.9994  | 0.9968  |
| Exactitud                     | n                                  | 9   | 8   |
|                               | Promedio de residuales (%)         | 0.10  | -0.05   |
|                               | Desviación Estándar                | 0.18  | 0.31  |
|                               | t <sub>exp</sub>                   | 1.662   | 0.459   |
|                               | t <sub>tab</sub> ( $\alpha=0.05$ ) | 2.365   | 2.365   |
| Repetibilidad                 | Promedios (%)                      | X <sub>1</sub> = 0.9  | X <sub>1</sub> = 3.1  |
|                               |                                    | X <sub>2</sub> = 5.1  | X <sub>2</sub> = 4.4  |
|                               |                                    | X <sub>3</sub> = 9.0  | X <sub>3</sub> = 6.8  |
|                               | Coefficientes de variación (%)     | CV <sub>1</sub> = 3.1   | CV <sub>1</sub> = 2.7   |
|                               |                                    | CV <sub>2</sub> = 1.1<br>CV <sub>3</sub> = 2.6<br>CV <sub>TOTAL</sub> = 2.3 | CV <sub>2</sub> = 4.4<br>CV <sub>3</sub> = 3.5<br>CV <sub>TOTAL</sub> = 4.0 |
| Precisión intermedia          | Promedio (%)                       | X = 5.0   | X = 4.4   |
|                               | Coefficiente de variación (%)      | CV = 3.6  | CV = 4.8  |
| Límite de Detección (LD)      | LD = 3·SD <sub>a</sub> / b         | 0.2%  | 0.2%  |
| Límite de Cuantificación (LQ) | LD = 10·SD <sub>a</sub> / b        | 0.8%  | 0.8%  |

Se evaluaron los parámetros de: selectividad, linealidad, exactitud, precisión, límites de detección y cuantificación; cumpliéndose para todos los parámetros los criterios establecidos.

A partir de las rectas de regresión obtenidas para la determinación del contenido de DH en las otras dos formas *pseudo*-polimórficas, se han estimado los límites de detección (LD) y de cuantificación (LQ) por espectroscopia NIR, obteniéndose en todos los casos valores inferiores a los obtenidos por difracción de rayos X.

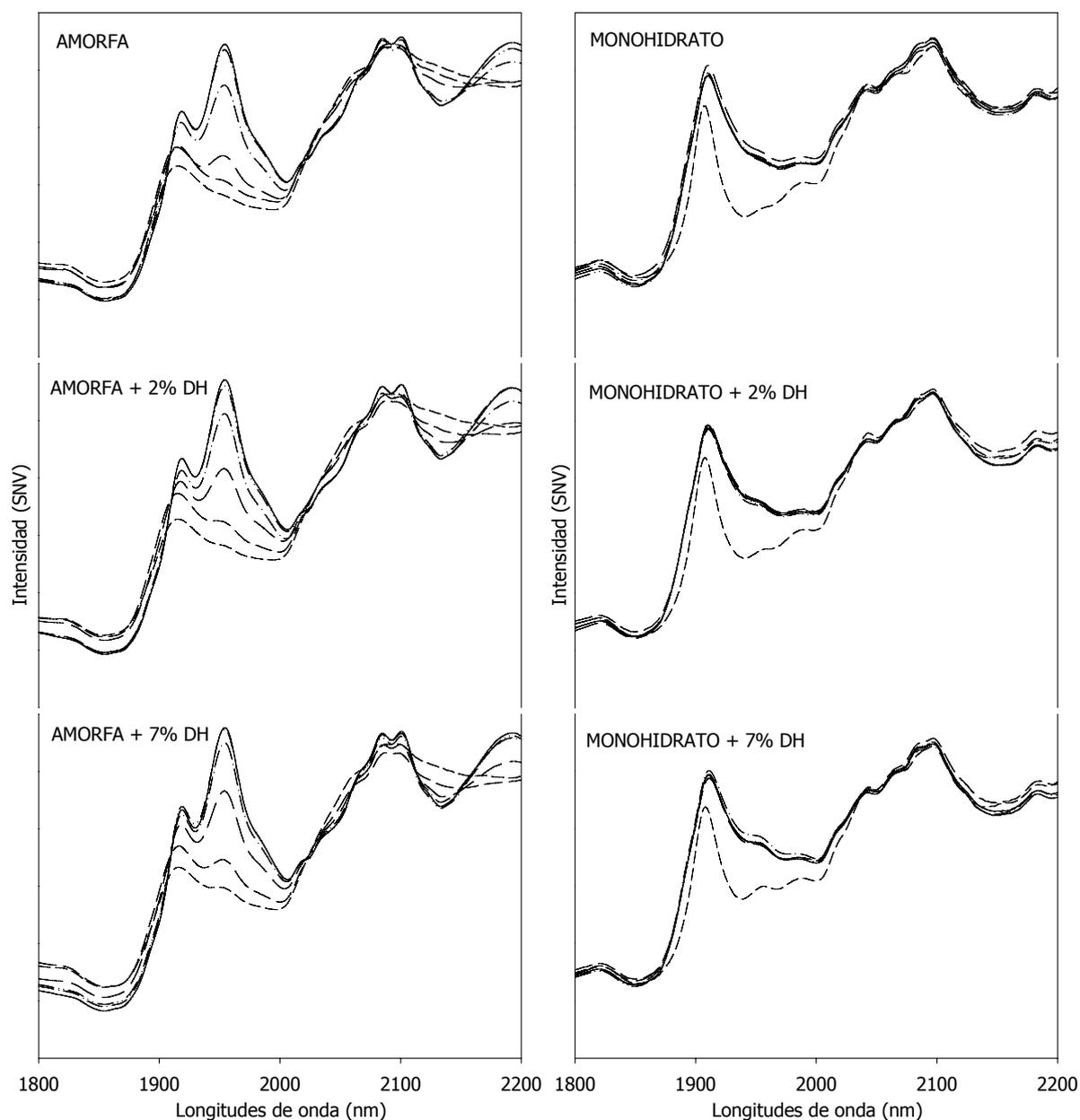
#### 6.1.3.2 Estudio de Estabilidad

Si bien es importante poder caracterizar los diferentes *pseudo*-polimorfos del fármaco, también es necesario poder detectar posibles transformaciones que pueden ocurrir en el proceso de producción, envasado y/o almacenamiento.

Se ha realizado un estudio de estabilidad de cada uno de los *pseudo*-polimorfos de la azitromicina en condiciones de elevada humedad y temperatura mediante espectroscopia NIR, con el objeto de establecer la mejor forma de conservación que asegure la existencia de una única forma polimórfica. Se estudia también la influencia de la presencia de DH en muestras de A y MH, sobre la estabilidad de las formas polimórficas.

Se prepararon muestras de A y MH con contenidos variables de DH (0%, 2% y 7%, respectivamente). Estas muestras se mantuvieron durante 60 días en dos condiciones de de humedad: en medio seco y saturado de humedad (100% HR) y

elevada temperatura (60°C). En la **Figura 16** se representa la evolución de los espectros NIR en la región 1800-2200nm de las formas A y MH puras y con contenidos variables de DH.



**Fig. 16** Evolución de los espectros NIR (SNV) de las formas A y MH puras y con contenidos variables de DH.

(1 día ---, 14 días - - -, 24 días . . ., 38 días -·-, 51 días -·-·-, 60 días -)

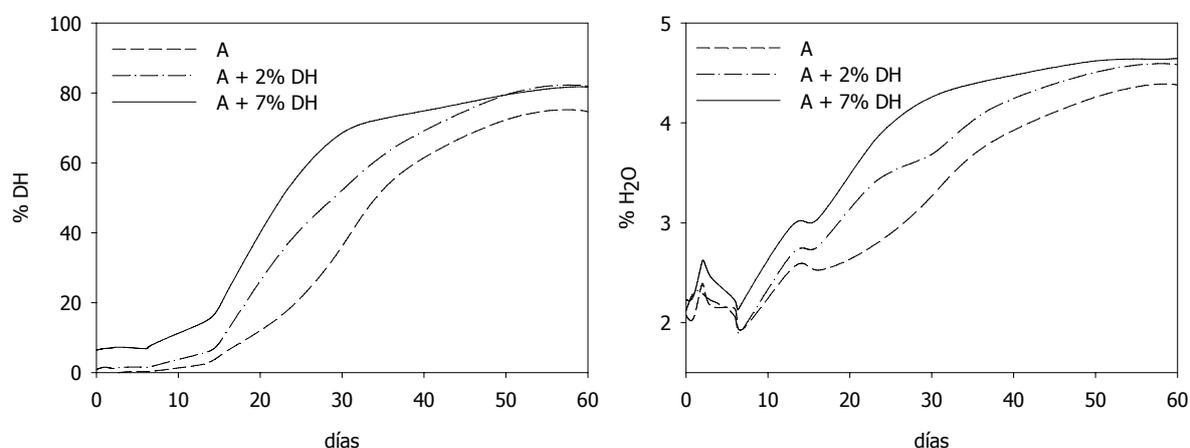
El análisis cualitativo de los espectros NIR muestra que cuando las formas *pseudo*-polimórficas (A, MH, DH) se mantienen en condiciones de elevada temperatura (60°C) no se producen transformaciones polimórficas. Sin embargo, el efecto combinado de humedad (100% de HR) y temperatura (60°C), no produce cambios en el DH, la cual puede ser considerada la forma más estable y provoca un comportamiento particular de las otras dos formas *pseudo*-polimórficas (A y MH). En estas mismas condiciones de humedad y temperatura, se estudió la influencia de diferentes contenidos de DH en la estabilidad de A y MH, ya que la presencia de cristales "madre" suele acelerar las transformaciones polimórficas.

En la **Figura 16** se aprecia que en las muestras de A se produce inicialmente un aumento en la intensidad de la banda del agua de adsorción (~1915 nm) y posteriormente empieza a aumentar la intensidad de la banda ~1955 nm que corresponde al agua de hidratación del DH, por tanto se ha producido la incorporación de agua en la estructura cristalina y por ende la transformación de la forma A en DH. La intensidad de la banda correspondiente a la formación del DH es mayor a medida que aumenta el contenido de DH en las muestras de A, lo que significa que la presencia de DH acelera la transformación polimórfica.

El comportamiento del MH difiere del de la A. En esta forma *pseudo*-polimórfica solamente se aprecia un aumento en la intensidad de la banda del agua en la primeras 24h, que después permanece invariable, lo que significa que ha adsorbido agua y no se ha producido transformación polimórfica en estas condiciones. La presencia de DH en MH, no altera la estabilidad de este *pseudo*-polimorfo.

A continuación discutiremos los resultados del análisis cuantitativo que se lleva a cabo empleando los modelos PLS1 desarrollados (Tabla 3). El modelo para la predicción del contenido de DH en A, tuvo que ser ajustado para expandir el rango de 0-100% DH.

En la Figura 17 se muestran las predicciones de DH y agua en las muestras de A.



**Fig. 17** Predicciones de los contenidos de DH y agua en muestras de A.

En la Figura 17 se aprecia que durante los primeros 14 días se produce un ligero aumento del contenido de agua, sin que se modifique el contenido de DH. Sin embargo, con el transcurso del tiempo el contenido de agua aumenta significativamente ( $\sim 2.5\%$ ), así como el de DH que llega a valores del 80%. Estos resultados cuantitativos confirman que se ha producido una transformación polimórfica de A en DH.

En las muestras de MH (0, 2, 7% de DH), solamente se produce un aumento del contenido de agua en las primeras 24h de aproximadamente 1.3%, que permanece prácticamente constante hasta el sexagésimo día. El contenido de DH no varía durante los 60 días del experimento, lo que significa que esta forma es estable en las condiciones establecidas.

El análisis cualitativo y cuantitativo demuestran que el MH y el DH son estables en las condiciones de humedad y temperatura establecidas, sin embargo la forma A se transforma en el DH que es la forma más estable.

## **6.2 Estudio y corrección del efecto de la temperatura en espectroscopia en el infrarrojo cercano**

### **6.2.1 Introducción**

La variación de la temperatura produce una variación de la intensidad de las bandas de absorción y un desplazamiento de la posición de las mismas en los espectros NIR de muestras líquidas.

Las características de las bandas de absorción, posición e intensidad, de los grupos funcionales OH, NH y SH dependen de las asociaciones intra- e intermoleculares, mientras que las características de las bandas de absorción de las vibraciones CH están fundamentalmente afectadas por los cambios de densidad.

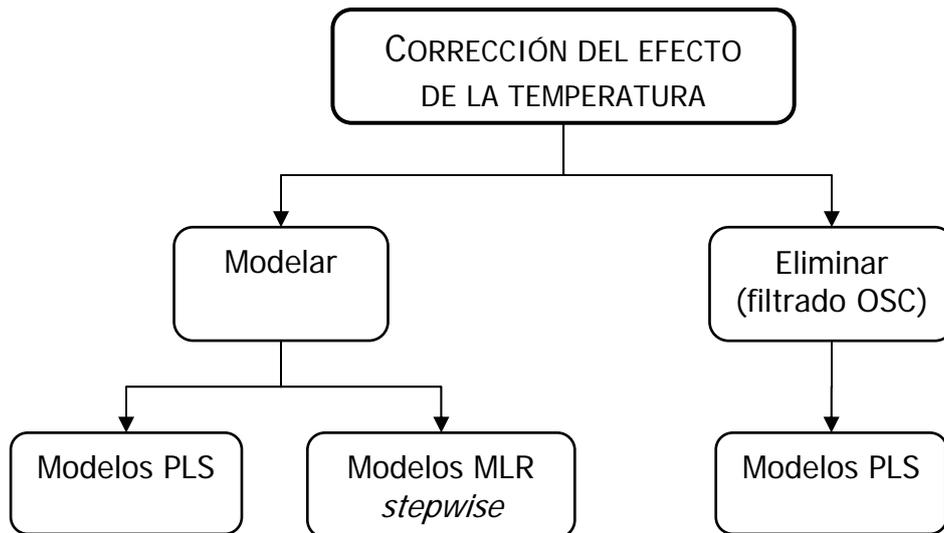
El cambio de temperatura ejerce un efecto más marcado en las bandas de los grupos funcionales que presentan enlaces por puentes de hidrógeno, ya que produce

una disminución del número promedio de puentes de hidrógeno por grupo OH y un aumento del número de OH libres.

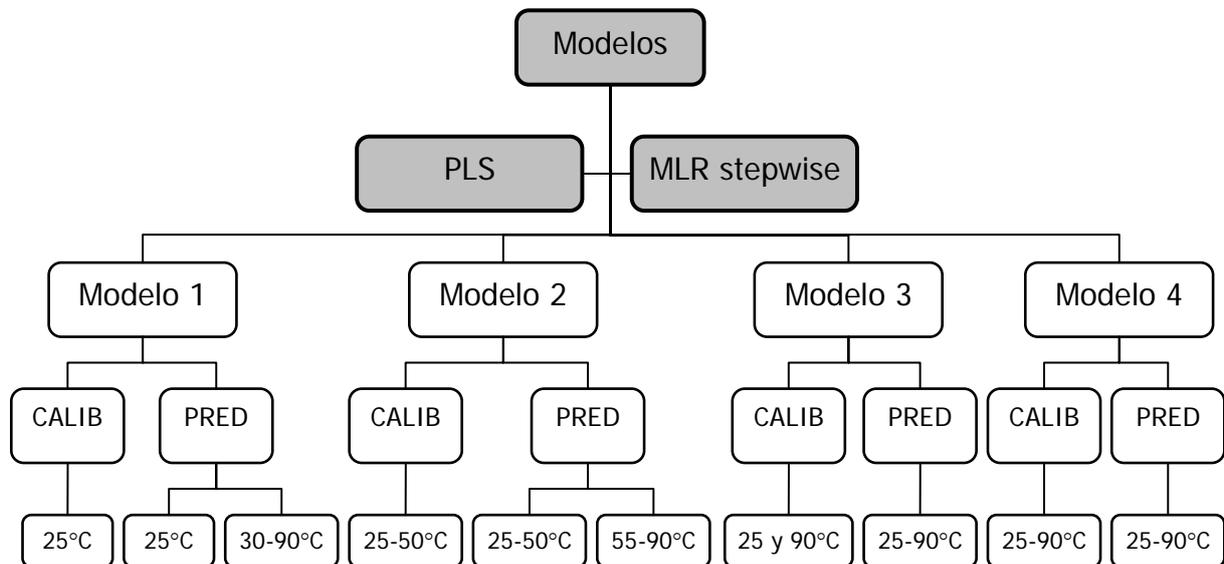
Estos cambios en las bandas de absorción, trae como consecuencia que la construcción de modelos de calibración con espectros registrados a una determinada temperatura o rango de temperaturas, no tiene una buena capacidad predictiva cuando se aplica a muestras que están registradas a temperaturas fuera del rango de las del modelo de calibración.

En la literatura se discute el efecto que tiene las variaciones de temperatura en la calibración multivariada y las diferentes metodologías aplicadas para la resolución de este problema. Los métodos propuestos se dividen en dos grupos, uno que intenta antes o durante la regresión de eliminar el efecto de la temperatura y otro que reúne a un conjunto de métodos de regresión "puros" donde se modela el efecto de la temperatura en vez de eliminarlo. Dentro del primer grupo se encuentra el algoritmo OSC que tiene la ventaja que no requiere conocer la temperatura a la cual han sido registradas las muestras y que discutiremos más adelante.

En esta memoria se lleva a cabo un estudio del efecto que la variación de la temperatura ejerce en la predicción de los productos implicados en la reacción de esterificación ( $\text{HAc} + \text{BuOH} \rightleftharpoons \text{BuAc} + \text{H}_2\text{O}$ ) en un intervalo de temperatura entre 25-90 °C. Por una parte se realiza un estudio del efecto que la temperatura ejerce sobre los espectros NIR de muestras líquidas y por otra parte se proponen diferentes estrategias para la corrección de este efecto en los modelos de calibración (**Figura 18**).



**Fig. 18** Estrategias para la corrección del efecto de la temperatura en los modelos de calibración.



**Fig. 19** Descripción de los conjuntos de calibración y predicción de los modelos PLS y MLR *stepwise*.

Como se muestra en la **Figura 18**, una estrategia seguida para la corrección del efecto de la temperatura es modelar el efecto de la temperatura y para ello se emplearon las regresiones PLS y MLR *stepwise*. En ambas regresiones se ensayaron diferentes intervalos de temperatura, de longitudes de onda y pretratamientos espectrales para conseguir modelos con la mejor capacidad predictiva en las diferentes situaciones experimentales.

En la **Figura 19** se presenta un esquema de los modelos ensayados con los respectivos conjuntos de calibración y predicción.

En el Modelo 1 el conjunto de calibración solamente contiene muestras registradas a 25°C y el conjunto de predicción se ha subdividido en dos: uno que contiene muestras a la misma temperatura que el conjunto de calibración y el otro subconjunto que contiene muestras del resto del intervalo de temperatura (30-90 °C). Para el Modelo 2 se construyó un conjunto de calibración con muestras del primer intervalo de temperatura (25-50 °C) y los dos conjuntos de predicción se construyeron con el mismo criterio del Modelo 1. El Modelo 3 tiene un conjunto de calibración que contiene muestras de los extremos del intervalo de temperatura (25 y 90 °C) y se predicen muestras en todo el rango de temperatura. El Modelo 4 contiene muestras en todo el intervalo de temperatura tanto en calibración como en predicción.

La otra estrategia seguida para la corrección del efecto de la temperatura es la utilización del filtrado de variables conocido como orthogonal signal correction (OSC) con el objetivo de eliminar la variabilidad espectral debida a la temperatura y no correlacionada con la concentración de los componentes implicados en la reacción de

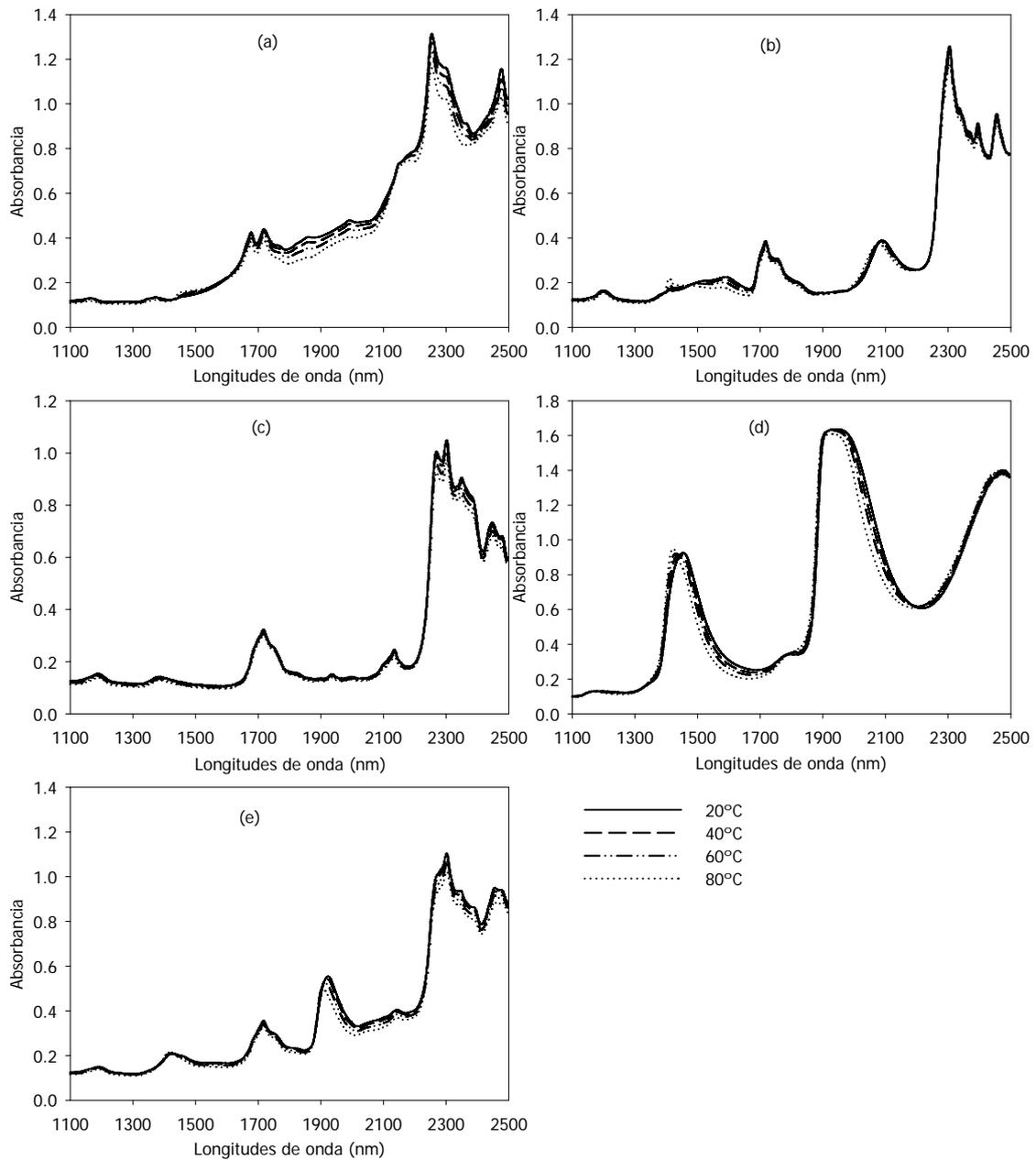
esterificación. Una vez eliminada la información espectral con el algoritmo OSC, se construyen modelos de calibración para la determinación de cada componente individual (PLS1) y simultáneamente de las cuatro especies implicadas (PLS2). También se realizó un estudio de la información espectral eliminada (matriz  $\mathbf{X}_E$ ) por el algoritmo OSC y de la información remanente (matriz  $\mathbf{X}_R$ ) que se utiliza en la construcción de los modelos de calibración.

### 6.2.2 Estudio del efecto de la temperatura sobre los espectros NIR

Se estudió el efecto que la temperatura ejerce sobre los espectros NIR de los compuestos puros implicados en la reacción de esterificación, así como de mezclas de estos (**Figura 20**).

El aumento de la temperatura provoca corrimientos hipsocrómicos de los máximos de absorbancia, sin cambios significativos en la intensidad de las bandas. El comportamiento de los cuatro compuestos puros ante el aumento de la temperatura difiere; presentando el espectro del agua las mayores alteraciones y el del acetato de butilo las menores. Los espectros del butanol y el ácido acético están más afectados que el acetato de butilo pero menos que el agua. El efecto de la temperatura en la mezcla cuaternaria (ver **Figura 20e**) depende de las bandas que se consideren: el desplazamiento es importante en las bandas de  $-\text{OH}$  y muy reducido en las de  $-\text{CH}$ . Podemos deducir que el desplazamiento más importante, unido a un estrechamiento de la banda, se produce en las bandas pertenecientes a los grupos susceptibles de formar enlace por puente de hidrógeno y este desplazamiento aumenta con la temperatura. Estos comportamientos indican que el efecto de la temperatura sobre

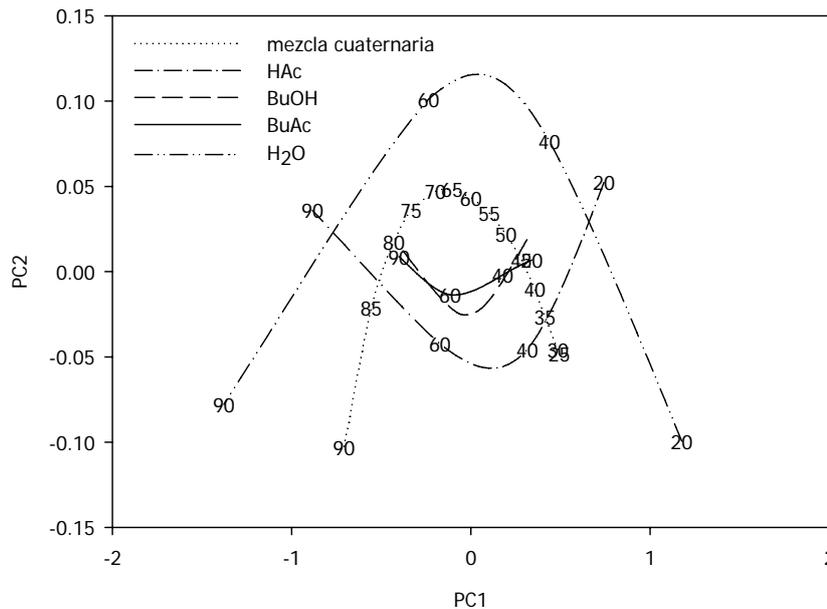
los espectros NIR, esta relacionado con la capacidad de los compuestos de formar enlaces por puente de hidrógeno y por tanto dependerá de la polaridad de los mismos.



**Fig. 20** Espectros NIR (SNV) registrados a diferentes temperaturas de:

(a) HAc, (b) BuOH, (c) BuAc, (d) H<sub>2</sub>O, (e) mezcla cuaternaria ( $X_{\text{HAc}}=0.26$ ,  $X_{\text{BuOH}}=0.24$ ,  $X_{\text{BuAc}}=0.25$ ,  $X_{\text{H}_2\text{O}}=0.25$ ).

Un análisis en componentes principales (PCA) permite apreciar las características de estas variaciones espectrales. En la **Figura 21** se muestra un gráfico de *scores* (PC2 vs. PC1) de los cuatro compuestos puros y una mezcla cuaternaria.



**Fig. 21** Gráfico de *scores* (PC2 vs. PC1) de HAc, BuOH, BuAc y H<sub>2</sub>O a 20, 40, 60, 90 °C y mezcla cuaternaria ( $X_{\text{HAc}}=0.26$ ,  $X_{\text{BuOH}}=0.24$ ,  $X_{\text{BuAc}}=0.25$ ,  $X_{\text{H}_2\text{O}}=0.25$ ) a 25-90 °C.

En el gráfico de los *scores* representado en la **Figura 21**, para todos los compuestos los dos primeros *scores* explican casi la totalidad de la varianza espectral (>99 %). Se obtiene una curva parabólica como resultado del aumento de la temperatura en todos los compuestos puros y la mezcla cuaternaria. Este mismo comportamiento se observa en todas las mezclas ternarias y cuaternarias. El H<sub>2</sub>O es el compuesto puro que muestra de forma más acusada esta tendencia, seguido del HAc, el BuOH y en menor grado el BuAc. La mezcla cuaternaria tiene un

comportamiento similar al del H<sub>2</sub>O, tanto en magnitud como en signo y es también más acusada en las mezclas con un alto contenido de H<sub>2</sub>O. Estos comportamientos indican claramente que los espectros NIR de los compuestos capaces de formar enlaces por puente de hidrógeno, están más afectados por el efecto de la temperatura.

### 6.2.3 Estrategias para la corrección del efecto de la temperatura

#### 6.2.3.1 Modelar el efecto de la temperatura

Con el objeto de modelar el efecto de la temperatura, se ensayaron modelos basados en el algoritmo de regresión parcial por mínimos cuadrados (PLS). La relación parabólica entre los *scores* de un PCA -como se mostró en la **Figura 19-**, sugirió que una regresión que incluyera términos cuadráticos debería mejorar la capacidad predictiva de los modelos. Por esta razón, se ensayaron también modelos mediante regresión con los *scores*, los cuadrados de los *scores* y combinaciones de los *scores* en pasos sucesivos (MLR *stepwise*).

En los modelos construidos con ambas regresiones, se ha incluido de manera implícita la variable temperatura en diferentes rangos -como se mostró en la **Figura 17-** y se ensayaron diferentes pre-tratamientos espectrales (derivadas y SNV), intervalos de longitudes de onda y factores.

En la **Tabla 5** se muestran los resultados de los modelos con ambas regresiones para cada uno de los analito.

**Tabla 5.** Parámetros de los modelos obtenidos con la regresiones PLS y MLR *stepwise*.

| Modelo | Analito          | Pretrat.                           | Regresión PLS |        |         |         | Regresión MLR <i>stepwise</i>   |        |         |         |
|--------|------------------|------------------------------------|---------------|--------|---------|---------|---|--------|---------|---------|
|        |                  |                                    | Fact.         | RSEC % | RSEP1 % | RSEP2 % | Fact.   | RSEC % | RSEP1 % | RSEP2 % |
| 1      | HAc              | 2 <sup>a</sup> D                   | 4             | 2.6    | 3.1     | 10.0    | PC1, PC3, PC2, PC1 <sup>2</sup> , PC5, PC4                                      | 2.7    | 3.4     | 10.4    |
| 1      | BuOH             | 1 <sup>a</sup> D                   | 4             | 3.0    | 3.6     | 10.7    | PC1, PC2, PC3, PC4, PC3 <sup>2</sup> , PC2 <sup>2</sup> , PC5                   | 2.2    | 3.4     | 7.3     |
| 1      | BuAc             | 2 <sup>a</sup> D                   | 5             | 4.1    | 4.6     | 7.4     | PC2, PC2 <sup>2</sup> , PC1, PC3, PC1 <sup>2</sup>                              | 4.1    | 5.0     | 8.1     |
| 1      | H <sub>2</sub> O | 1 <sup>a</sup> D                   | 4             | 5.1    | 5.1     | 15.9    | PC2, PC1, PC3, PC1 <sup>2</sup> , PC5   | 5.0    | 6.2     | 8.4     |
| 2      | HAc              | 2 <sup>a</sup> D                   | 5             | 3.2    | 3.9     | 9.4     | PC1, PC2, PC3, PC5, PC2 <sup>2</sup> , PC4, PC1 <sup>2</sup>                    | 4.2    | 4.9     | 10.5    |
| 2      | BuOH             | 1 <sup>a</sup> D                   | 5             | 3.0    | 3.9     | 9.2     | PC1, PC3, PC2, PC5, PC3 <sup>2</sup> , PC1 <sup>2</sup> , PC2 <sup>2</sup>      | 3.1    | 3.5     | 6.7     |
| 2      | BuAc             | 2 <sup>a</sup> D                   | 5             | 4.2    | 4.2     | 5.5     | PC2, PC1, PC2 <sup>2</sup> , PC1 <sup>2</sup>                                   | 5.3    | 5.3     | 6.8     |
| 2      | H <sub>2</sub> O | 2 <sup>a</sup> D/1 <sup>a</sup> D* | 5             | 4.3    | 4.3     | 7.8     | PC2, PC1, PC3, PC1 <sup>2</sup> , PC3 <sup>2</sup> , PC4, PC5, PC2 <sup>2</sup> | 3.6    | 4.0     | 6.8     |
| 3      | HAc              | 2 <sup>a</sup> D                   | 5             | 5.0    | 7.7     | -       | PC1, PC2, PC3, PC4, PC5, PC2 <sup>2</sup> , PC1 <sup>2</sup>                    | 5.7    | 7.4     | -       |
| 3      | BuOH             | 1 <sup>a</sup> D                   | 5             | 3.7    | 4.6     | -       | PC1, PC2, PC3, PC5, PC1 <sup>2</sup> , PC5 <sup>2</sup>                         | 5.5    | 7.3     | -       |
| 3      | BuAc             | 2 <sup>a</sup> D                   | 5             | 6.3    | 6.7     | -       | PC2, PC1, PC2 <sup>2</sup> , PC4, PC1 <sup>2</sup> , PC3                        | 5.5    | 5.4     | -       |
| 3      | H <sub>2</sub> O | 1 <sup>a</sup> D                   | 5             | 5.8    | 12.4    | -       | PC2, PC1, PC3, PC1 <sup>2</sup> , PC4, PC5, PC3 <sup>2</sup>                    | 6.6    | 8.9     | -       |
| 4      | HAc              | 2 <sup>a</sup> D                   | 5             | 5.3    | 5.7     | -       | PC2, PC1, PC3, PC4, PC5, PC2 <sup>2</sup> , PC4 <sup>2</sup>                    | 6.3    | 6.2     | -       |
| 4      | BuOH             | 1 <sup>a</sup> D                   | 5             | 3.7    | 3.6     | -       | PC1, PC3, PC5, PC2, PC2 <sup>2</sup>  | 5.6    | 5.4     | -       |
| 4      | BuAc             | 2 <sup>a</sup> D                   | 5             | 4.7    | 4.7     | -       | PC2, PC1, PC2 <sup>2</sup> , PC4, PC1 <sup>2</sup> , PC3                        | 5.6    | 5.7     | -       |
| 4      | H <sub>2</sub> O | 1 <sup>a</sup> D                   | 5             | 6.0    | 5.8     | -       | PC2, PC3, PC1, PC2 <sup>2</sup> , PC5, PC4 <sup>2</sup>                         | 5.4    | 5.1     | -       |

Pretrat. – Pretratamiento Espectral, Fact. – Factores,

\* Es el único caso donde el pretratamiento espectral que daba los mejores resultados no concordaba en las dos regresiones. En el modelo PLS se utilizó la 2<sup>a</sup>D y en el MLR *stepwise* la 1<sup>a</sup>D.

Todos los modelos fueron construidos en el intervalo espectral 1100-2500 nm, que arrojó los mejores resultados.

Como se puede apreciar en la **Tabla 5**, los mejores resultados se obtienen con espectros de derivadas en el intervalo de 1100-2500 nm.

De forma general en los modelos PLS, la inclusión de muestras con diferentes temperaturas en los modelos de calibración, aumenta en uno el número de factores PLS.

En los modelos MLR *stepwise*, generalmente la selección de factores sigue el siguiente orden: primero los *scores*, después los cuadrados de los *scores* y en ningún caso los términos de combinaciones de *scores*. No se observa una clara relación entre la incorporación de la variable temperatura en los modelos y la inclusión de términos cuadráticos en el modelo. Los modelos con espectros de una sola temperatura incluyen también términos cuadráticos.

Para todos los compuestos y en ambas regresiones, el error de predicción de los conjuntos de muestras registradas a temperaturas no incluidas en las calibraciones son mayores que las que están en el mismo intervalo de temperatura que las calibraciones ( $RSEP2 > RSEP1$ ).

Los modelos (PLS y MLR *stepwise*) del BuAc, que incluyen de forma parcial la variabilidad de la temperatura (Modelo 2 y Modelo 3), logran predecir correctamente el contenido de este analito en muestras registradas en todo el rango de temperatura. Este comportamiento singular del BuAc se debe a que es el analito menos influenciado por la temperatura, debido a la imposibilidad de formar enlaces por puente de hidrogeno.

Para el resto de los analitos es necesario construir modelos que contienen la variabilidad de la temperatura en todo el rango (Modelo 4), para lograr predicciones satisfactorias. Una excepción la constituye el Modelo 3 basado en la regresión PLS, que permite predecir correctamente el contenido de BuOH.

Estos resultados concuerdan con lo observado en los gráficos de scores de la Figura 2, que indica que el H<sub>2</sub>O y el HAc son los analitos más influenciados por la temperatura y el BuAc y el BuOH son los menos afectados.

Los resultados de los modelos de calibración PLS1 y MLR stepwise son similares y deben incluir parcial o totalmente el efecto de la temperatura para lograr predicciones correctas.

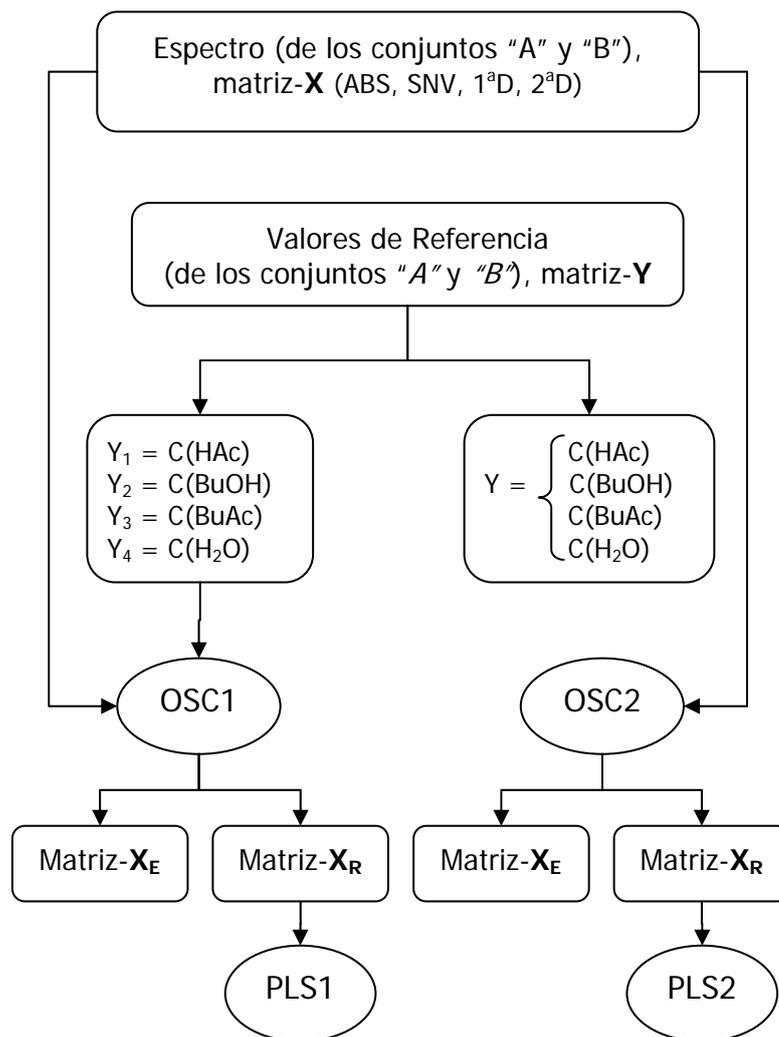
#### **6.2.3.2 Eliminar el efecto de la temperatura**

La segunda estrategia para la corrección del efecto de la temperatura, consiste en eliminar o reducir el efecto de la temperatura, mediante el filtrado de variables denominado *Orthogonal Signal Correction* (OSC). Este pre-tratamiento espectral permite eliminar la variabilidad espectral no correlacionada con la concentración de los analitos, como es la temperatura. Con la información espectral que esta correlacionada con la concentración de los analitos y no con la temperatura, se construyen modelos de calibración mediante la regresión PLS.

Se han preparado dos modelos: uno constituido por muestras de calibración registradas a una sola temperatura (Conjunto "A") y el otro contiene muestras que

cubren el rango completo de temperaturas (Conjunto "B"). El resto de las muestras no usadas en la calibración se utilizan en el set de predicción.

En la **Figura 22** se representa el esquema seguido en el estudio, indicando las etapas de la construcción de los modelos de calibración PLS1 y PLS2, después de aplicar el pre-tratamiento espectral OSC.



**Fig. 22** Esquema seguido en la construcción de los modelos de calibración mediante el algoritmo OSC.

$C(\text{analito})$  – Fracción Molar de los analitos, matriz  $\mathbf{X}_E$  – información espectral que es ortogonal a la matriz  $\mathbf{Y}$  y que es eliminada con el OSC, matriz  $\mathbf{X}_R$  – información espectral que queda remanente después de aplicar el OSC.

A los espectros del conjunto de calibración (matriz  $\mathbf{X}$ ) se les aplica el OSC y se extrae la información espectral ortogonal a la matriz  $\mathbf{Y}$  (las matrices  $\mathbf{Y}_1$ ,  $\mathbf{Y}_2$ ,  $\mathbf{Y}_3$ ,  $\mathbf{Y}_4$  contienen las fracciones molares de cada uno de los analitos o de todos los analitos, matriz  $\mathbf{Y}$ ) de los datos analíticos (matriz  $\mathbf{X}_E$ ); la información espectral resultante de la eliminación (matriz  $\mathbf{X}_R$ ) se utiliza en la construcción de los modelos PLS. En cada caso, con la matriz resultante (matriz  $\mathbf{X}_R$ ) se construyen los cuatro modelos basados en la regresión PLS1 para la determinación de cada uno de los cuatro analitos o el modelo PLS2 para la determinación simultánea de los cuatro analitos. Estas dos estrategias de trabajo se aplicaron a los conjuntos de muestras "A" y "B".

Al aplicar el algoritmo OSC a las muestras del conjunto "A" se obtienen, tanto con la regresión PLS1 como con la PLS2, modelos con buena capacidad predictiva en muestras registradas a 25°C (RSEP < 5%), sin embargo la predicción de muestras registradas a temperaturas distintas a las de la calibración conduce a errores importantes en todos los casos (RSEP > 10%). Sin embargo, la aplicación al conjunto "B", que contiene muestras registradas a todas las temperaturas, conduce a modelos PLS1 simples y con una buena capacidad predictiva. En la **Tabla 6** se muestran las características de los modelos construidos con el conjunto "B".

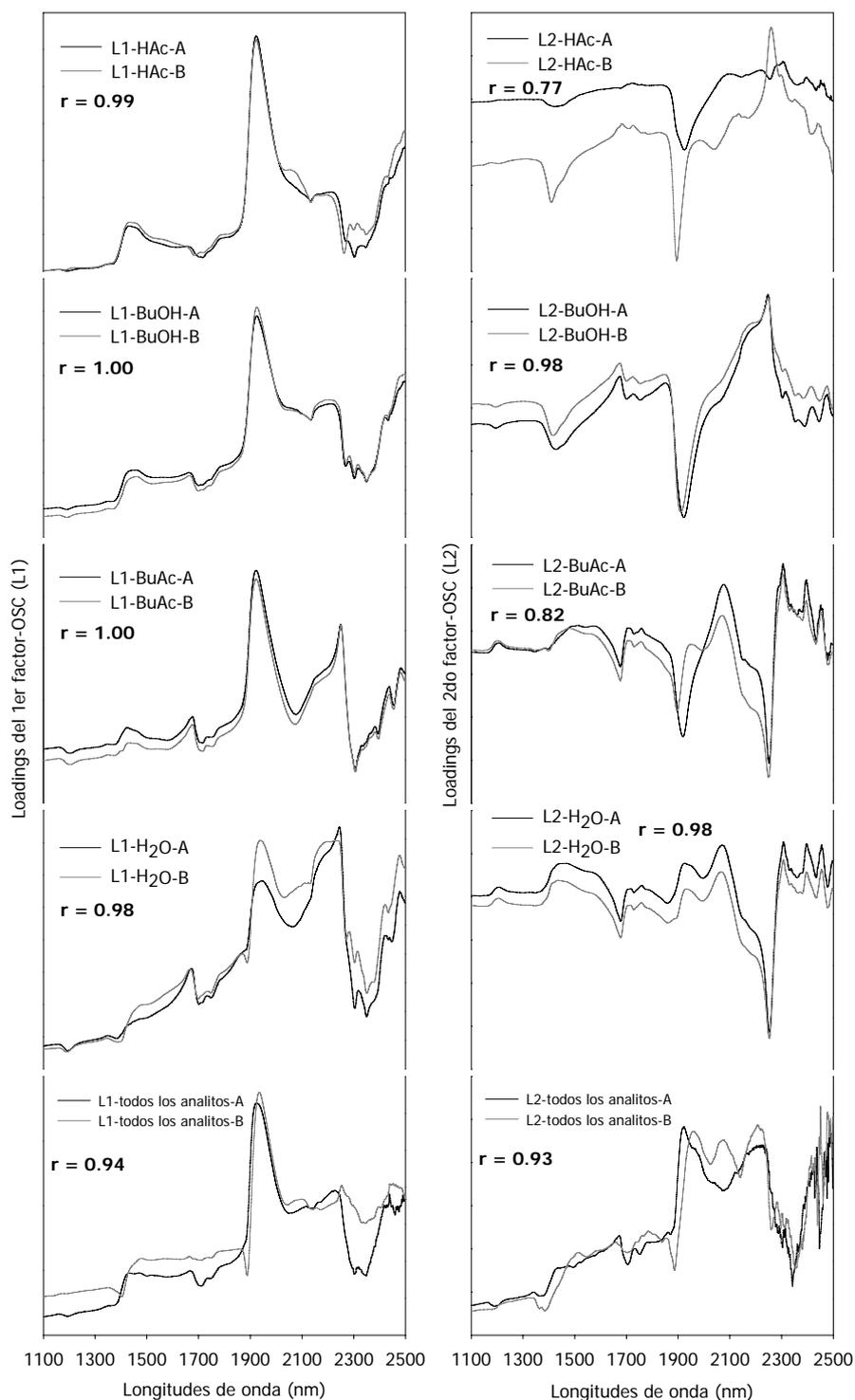
**Tabla 6.** Características del Modelo "B" para los cuatro analitos.

|      |                | HAc | BuOH             | BuAc | H <sub>2</sub> O |
|------|----------------|-----|------------------|------|------------------|
| PLS1 | Modo Espectral | ABS | 2 <sup>a</sup> D | ABS  | ABS              |
|      | Factores OSC   | 2   | 1                | 2    | 2                |
|      | Factores PLS   | 1   | 1                | 1    | 1                |
|      | RSEC (%)       | 2.0 | 2.3              | 1.8  | 3.2              |
|      | RSEP (%)       | 3.2 | 3.7              | 2.5  | 4.3              |
| PLS2 | Modo Espectral | ABS | 2 <sup>a</sup> D | ABS  | ABS              |
|      | Factores OSC   | 2   | 2                | 2    | 2                |
|      | Factores PLS   | 3   | 3                | 3    | 3                |
|      | RSEC (%)       | 6.7 | 2.1              | 5.1  | 5.6              |
|      | RSEP (%)       | 7.0 | 2.5              | 6.6  | 6.2              |

Todos los modelos fueron construidos en el intervalo espectral 1100-2500 nm, que arrojó los mejores resultados.

El número de factores OSC necesarios para eliminar la información no correlacionada es el mismo en los dos pretratamientos (2 factores OSC). Los modelos PLS1 requieren un menor número de factores PLS que los modelos PLS2 y también la calidad de los modelos (PLS1) tanto en calibración como en predicción es mejor (**Tabla 6**). Esto demuestra la mayor efectividad del OSC1 frente al OSC2 en la eliminación de la información espectral no correlacionada con la concentración.

Complementariamente a la aplicación del algoritmo OSC con el objetivo de eliminar la variabilidad espectral debida al efecto de la temperatura, se llevó a cabo un estudio para conocer el tipo de información espectral eliminada por la aplicación del tratamiento OSC (matriz  $\mathbf{X}_E$ ) y la que queda remanente (matriz  $\mathbf{X}_R$ ) con la cual se construyen los modelos de calibración PLS1 y PLS2, utilizando para ello los *loadings* OSC y los *scores* del modelo PCA.



**Fig. 23** Loadings del primer y segundo factor-OSC tomando como matriz-**X** las muestras de los conjuntos "A" y "B" y como matriz-**Y** las fracciones molares de cada uno de los analitos y de todos a la vez.

En la **Figura 23** se muestran los *loadings* de los factores-OSC (L1, L2) para los conjuntos "A" y "B". Como se puede apreciar, los *loadings* del primer factor-OSC (L1) para los modelos corregidos de los cuatro analitos son muy similares en los dos conjuntos "A" y "B" (los coeficientes de correlación son superiores a 0.98). La similitud entre los *loadings* del primer factor-OSC en ambos modelos indica que este factor-OSC no elimina el efecto de la temperatura, ya que el conjunto "A" solo contiene espectros registrados a la misma temperatura. Ambos modelos eliminan el mismo tipo de contribuciones espectrales y para todos los analitos la información eliminada se asemejan al espectro de agua que es el analito con la contribución mas importante al espectro de las mezclas. Cuando la corrección OSC se realiza usando la matriz **Y** (que contiene las fracciones molares de todos los analitos) el coeficiente de correlación entre ambos *loadings* es algo menor ( $r=0.94$ ), lo que indica que el primer *loading* del conjunto "B" está corrigiendo en alguna medida el efecto de la temperatura.

Los *loadings* del segundo factor-OSC (L2) de los conjuntos "A" y "B", tienen una menor similitud, con coeficientes de correlación menores (0.77 para el HAc y 0.82 para el BuAc), lo que indica que este segundo factor corrige en mayor medida el efecto de la temperatura. Los perfiles de los *loadings* presentan un perfil muy semejante pero con desplazamientos de los picos que es el principal efecto que produce el cambio de temperatura sobre los espectros NIR y que demuestran la necesidad de un segundo factor-OSC para corregir el efecto de la temperatura.

Estos hechos demuestran que la principal causa de variación entre los espectros no es debido a la variación de temperatura. Los segundos *loadings* de la

corrección OSC2 con ambos conjuntos A y B son más diferentes entre sí, lo que es indicativo de la importante corrección que produce el segundo factor OSC en el modelo del conjunto B.

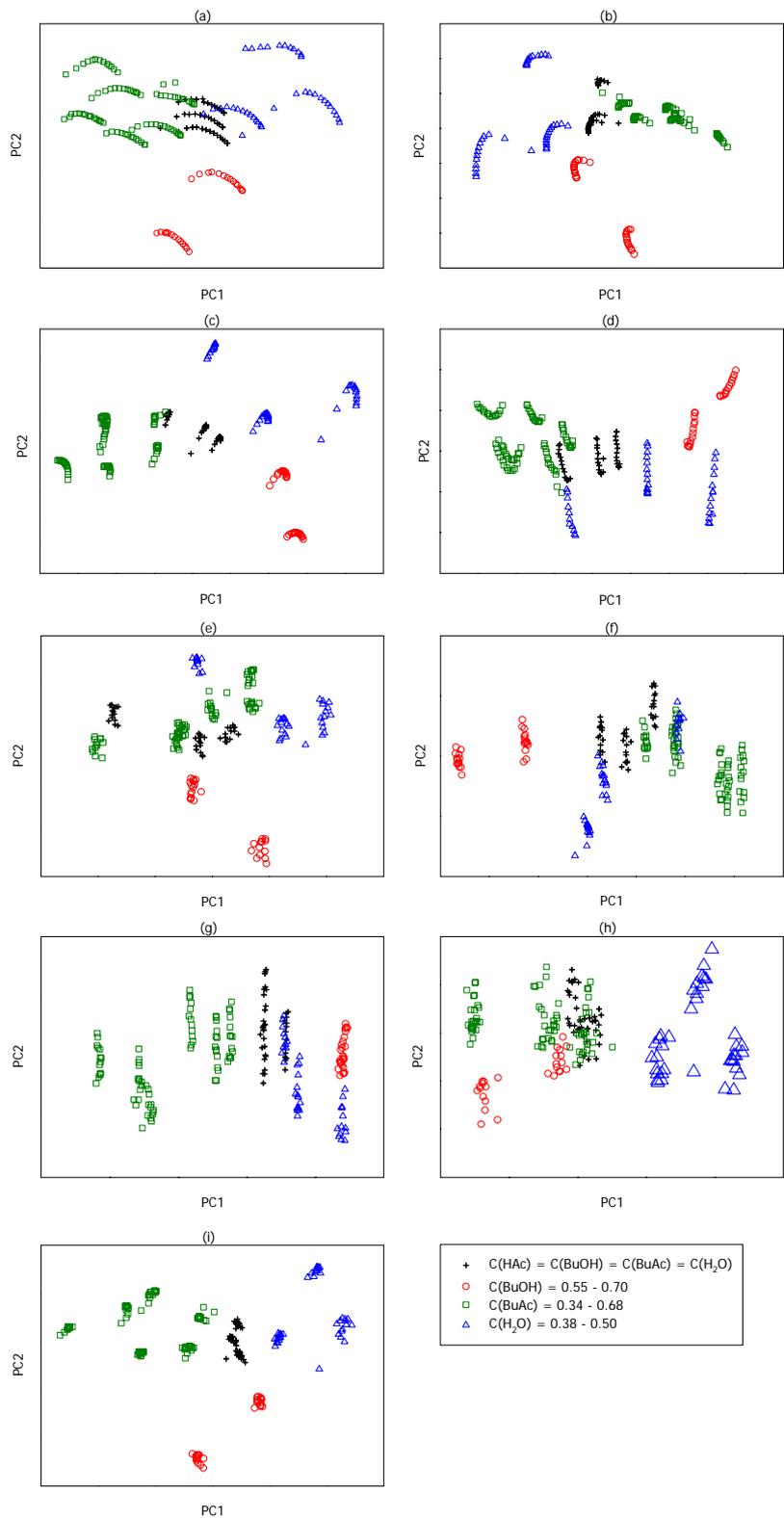
Las características de la información espectral remanente (matriz  $\mathbf{X}_R$ ) después de aplicar la corrección de dos factores OSC, puede observarse comparando los gráficos de dispersión de scores (PC2 vs. PC1) de los espectros antes y después de aplicar el algoritmo OSC a un conjunto de muestras registradas en todo el rango de temperatura (25-90°C). En la **Figura 24** se muestran los diferentes gráficos de dispersión de scores PC2 vs. PC1 del PCA realizado con la matriz original  $\mathbf{X}$  en los modos Absorbancia, SNV, 1aD ó 2aD y las matrices  $\mathbf{X}_R$  resultantes de aplicar el OSC1 y OSC2.

Un PCA de la matriz  $\mathbf{X}$  en modo absorbancia da lugar a una variación parabólica de los dos scores debido a la variación de la temperatura en todas las mezclas, más o menos acusada según su composición (ver **Figura 24a**). Al aplicar los pre-tratamientos de SNV y derivadas (1aD y 2aD) se mantiene de modo general esta relación parabólica, siendo en algunas muestras más acentuada que en otras (ver **Figuras 24 b, c, d**).

Al aplicar el algoritmo OSC con dos factores-OSC, se obtiene una tendencia de los *scores* claramente diferente. El gráfico de dispersión del PCA de la matriz  $\mathbf{X}_R$  (tomando la matriz  $\mathbf{Y}$  con la fracción molar de cada uno de los analitos, **Figuras 24 e, f, g, h**) indica que el valor del primer componente PC1 está directamente relacionado con la composición de la muestra (fracción molar del analito que ha sido seleccionado como matriz Y), mientras que el desplazamiento debido al segundo

componente PC2 está asociado con la variación de la temperatura; esto demuestra que el OSC no elimina completamente su efecto. Sin embargo, el hecho de que el PC1 esté directamente relacionado con la fracción molar de los analitos, permite la construcción de modelos de calibración muy simples (1 factor-PLS) con una adecuada capacidad predictiva.

Después de aplicar un OSC2 con dos factores, el PCA de la matriz remanente  $X_R$  (la matriz Y contiene las fracciones molares de los cuatro analitos) desaparece esta variación de los *scores* y las muestras de composición semejante y registradas a diferentes temperaturas aparecen agrupadas en clusters compactos (**Figura 21i**) según su composición, lo que demuestra que se ha eliminado la variabilidad espectral debida a la variación de temperatura. La posición de las mezclas en el espacio bidimensional de los scores es la misma que antes de aplicar el algoritmo OSC2 (ver **Figuras 24a y 24i**), lo cual indica que solamente se ha eliminado la variabilidad espectral debida a la temperatura.



**Fig. 24** Gráficos *scores* (PC2 vs. PC1) del PCA con las variables en modo:

- (a) Absorbancia, (b) SNV, (c) 1aD, (d) 2aD, (e) OSC1-HAc, (f) OSC1-BuOH, (g) OSC1-BuAc, (h) OSC1-H<sub>2</sub>O, (i) OSC2.



## CONCLUSIONES

## CONCLUSIONES

Del estudio de la influencia de parámetros físicos en el desarrollo de los modelos de calibración, se pueden extraer las siguientes conclusiones:

1. Se demuestra que la espectroscopia NIR es una técnica adecuada para la caracterización de tres formas polimórficas de la azitromicina y que presenta indudables ventajas sobre la difracción de rayos X en rapidez y sensibilidad, lo que nos permite proponerla como técnica alternativa en el control analítico.
2. La caracterización de cada una de las diferentes formas polimórficas se ha realizado a través de la construcción de bibliotecas que también permiten fijar la calidad mínima necesaria para la utilización adecuada de cada una de las diferentes formas.
3. Los modelos desarrollados mediante el algoritmo PLS para la cuantificación del contenido de dihidrato en muestras de amorfa y monohidrato poseen características de exactitud y precisión adecuadas para su aplicación en el control de proceso en industria farmacéutica. Previamente los métodos se han validado siguiendo las guías propuestas. También se han desarrollado modelos PLS para la determinación del contenido total de agua que puede servir como indicativo de la forma polimórfica presente.
4. El análisis cualitativo y cuantitativo de muestras mantenidas en condiciones de temperatura y humedad de transformación acelerada indica que las formas cristalinas (monohidrato y dihidrato) son estables, en cambio la forma amorfa es ligeramente higroscópica y se transforma lentamente en el dihidrato.

5. El aumento de la temperatura produce en los espectros NIR de muestras líquidas un desplazamiento hipsocrómico de las bandas de absorción y un estrechamiento de las mismas. Este desplazamiento es importante en las bandas de combinación de  $-OH$  y muy reducido en las de  $-CH$  y es menos acusado en los bandas de sobretodo.
6. Un insuficiente control de la temperatura durante el registro de los espectros produce un aumento de la complejidad de los modelos de calibración y un importante efecto en la capacidad predictiva de los modelos, siendo este efecto mayor en compuestos polares que en los no polares.
7. La disminución de la capacidad predictiva causada por la temperatura se puede corregir mediante dos estrategias diferentes que se basan en el modelado (inclusión en el conjunto de calibración de espectros de muestras registradas a las diferentes temperaturas) o en la eliminación del efecto de la misma (aplicación de un algoritmo que elimine la información espectral relacionada con la temperatura).
8. Los algoritmos PLS y MLR *stepwise* aplicados a la estrategia de modelado producen resultados similares y las regresiones deben incluir total o parcialmente el efecto de la temperatura para lograr predicciones correctas. Los modelos resultantes son bastante complejos.
9. La estrategia de eliminación de la información espectral no correlacionada con la temperatura se ha aplicado mediante el algoritmo OSC. El filtrado de variables OSC reduce en gran medida la información espectral relacionada con la temperatura, a diferencia de otros pre-tratamientos (derivadas y SNV). Para que

la eliminación sea efectiva es necesario que los modelos construidos con el algoritmo OSC incluyan la variabilidad debida a la temperatura y son necesarios dos factores OSC para la eliminación. Los modelos PLS construidos después de la eliminación son más simples y poseen una mejor capacidad predictiva, que los de modelado.

10. Los *loadings* de los modelos OSC muestran las características de la información espectral eliminada y los *scores* de los modelos PCA la información remanente después de aplicar el algoritmo OSC. El análisis de los *loadings* demuestra la necesidad de dos factores OSC para corregir el efecto de la temperatura, mientras que un solo factor PLS es necesario para obtener una buena capacidad predictiva en los modelos PLS.

ANEXOS

# ANEXO I

## “CHARACTERIZATION AND ANALYSIS OF POLYMORPHS BY NEAR-INFRARED SPECTROMETRY”

*M. Blanco, D. Valdés, M.S. Bayod, F. Fernández-Martí, I. Llorente*

Analytica Chimica Acta Vol. 502, No. 2 (**2004**) 221-227.

## ANEXO II

### “APPLICATION OF NIR SPECTROSCOPY IN POLYMORPHIC ANALYSIS: STUDY OF *PSEUDO*- POLYMORPHS STABILITY”

*M. Blanco, D. Valdés, I. Llorente, M.S. Bayod*

Journal of Pharmaceutical Sciences Vol. 94, No. 6 (2005) 1336-1342.

# ANEXO III

“INFLUENCE OF TEMPERATURE ON THE  
PREDICTIVE ABILITY OF NEAR INFRARED  
SPECTROSCOPY MODELS”

*M. Blanco and D. Valdés*

J. Near Infrared Spectrosc. Vol. 12, No. 2 (2004) 121-126.

# ANEXO IV

“SUPPRESSING THE TEMPERATURE EFFECT IN  
NEAR INFRARED SPECTROSCOPY BY USING  
ORTHOGONAL SIGNAL CORRECTION”

*M. Blanco and D. Valdés*

Enviado a publicar