



**Universitat Autònoma  
de Barcelona**

# Validación y transferencia de métodos analíticos mediante espectroscopia NIR

---

Construcción y validación de modelos quimiométricos para  
la identificación y cuantificación de productos  
farmacéuticos.

**Cristina Monfort Fraga**

Junio 2014

Directora (UAB): Dra. Maria Muñoz Tapia

Directora (KernPharma): Dra. Anna Peguero Gutiérrez.

Trabajo de Fin de Grado presentado a la Facultad de Ciencias

de la Universitat Autònoma de Barcelona para la obtención del Grado en Química



*“Si buscas resultados distintos, no hagas siempre lo mismo.”*

Albert Einstein



## AGRADECIMIENTOS

Este trabajo de fin de grado me ha permitido enfrentarme a retos y responsabilidades debido a que su realización se basa en el trabajo desarrollado en una de las más renombradas empresas farmacéuticas españolas, KernPharma.

Debo agradecer, por tanto, por partida doble, tanto a la Universidad Autónoma de Barcelona por ofrecerme la posibilidad de tener un primer contacto con el mundo laboral como a la propia empresa por ofrecerme un puesto desde el que he podido poner en práctica todo lo que desde la facultad he aprendido.

Muy especialmente me gustaría agradecer la realización de este trabajo a la doctora Anna Peguero Gutiérrez, mi tutora en la empresa, por haberme dedicado tiempo y esfuerzo, por haberme resuelto todo tipo de dudas, por enseñarme una ética de trabajo constante y disciplinada y por último, por predicarme que la curiosidad es lo último que un químico debe perder.

Dedico este trabajo a mi abuelo, Jesús, fallecido el año pasado, por enseñarme y recordarme con cariño en tiempos de estrés que “Quién hace lo que puede no está obligado a más”.

Sinceramente, gracias.

Cristina.



# ÍNDICE

índice de abreviaturas .....	1
Índice de figuras .....	1
Resumen .....	4
Abstract .....	4
1. Objetivos .....	5
2. Introducción a los métodos espectroscópicos .....	6
2.1. Absorción de la radiación .....	6
2.2. Requisitos para la absorción en la región del infrarrojo .....	7
2.3. Fundamento de la absorción en NIR .....	7
2.4. Instrumentación para NIR .....	7
2.5. Funcionamiento del FT-NIR de Bruker: .....	8
3. Quimiometría .....	9
3.1. Análisis Multivariable .....	9
3.1.1. Método cualitativo: creación de una biblioteca espectral .....	10
3.2. Análisis de Clusters (o conglomerados) .....	12
3.3. Análisis en componentes principales (PCA) .....	12
3.3.1 Método cuantitativo: Uso de Análisis de Regresión y modelos predictivos .....	15
3.4. Métodos de regresión multivariable .....	16
3.5. Validación de métodos analíticos .....	17
4. Metodología .....	18
4.1. Descripción de la creación del modelo de identificación .....	18
4.2. Descripción de la creación del modelo de cuantificación .....	19
5. Procedimiento experimental .....	22
5.1. Instrumentación .....	22
5.2. Registro de espectros para el modelo de identificación .....	23
5.3. Preparación de las muestras de laboratorio .....	24
5.4. Método de referencia: HPLC .....	25
6. Resultados y discusión .....	27
6.1. Modelo de identificación .....	27
6.1.1. Especialidad farmacéutica en comprimidos recubiertos .....	27
6.1.2. Especialidad farmacéutica en forma de granulado efervescente .....	30
6.2. Modelo de cuantificación .....	31
6.2.1. Especialidad farmacéutica en comprimidos recubiertos .....	31
6.2.2. Especialidad farmacéutica en forma de granulado efervescente .....	38
7. Conclusiones .....	41
8. Bibliografía .....	42

## ÍNDICE DE ABREVIATURAS

NIRS	Near Infrared spectroscopy
API	Active Pharmaceutical Ingredient
LEDs	Light-Emitting diodes
AOTF	Acousto-Optic Tunable Filter
FT-NIR	Fourier Transform Near Infrared
CCDs	Charge-Coupled Devices
MVA	Multivariate Analysis
SNV	Standard Normal Variate
1º der, 2º der	1º derivative, 2º derivative
PCA	Principle Component Analysis
PC1, PC2	Principle Component 1, Principle Component 2
MLR	Multiple Linear Regression
PCR	Principle Component Regression
PLSR	Partial Least Squares Regression
RMSEP	Root Mean Square Error of Prediction
RPD	Residual Prediction Deviation
R <sup>2</sup>	Correlation Coefficient
HPLC	High Pressure Liquid Chromatography
EMA	European Medicines Agency
SEL <sub>HPLC</sub>	Standard Error of Laboratory (HPLC)
SEP	Standard Error of Prediction
RSD	Relative Standard Deviation
SD	Standard Deviation



## ÍNDICE DE FIGURAS

### Tablas

Tabla 1: Procedimientos para demostrar los parámetros de calidad del modelo, según la Agencia Europea del Medicamento (EMA) .....	21
Tabla 2: Preparación de las muestras patrón en un rango de concentración del 70-130%.....	24
Tabla 3: Concentración de los excipientes del núcleo. ....	24
Tabla 4: Parámetros espectrales de la biblioteca de identificación. ....	27
Tabla 5: Umbrales y niveles de confianza característicos de cada clase para la identificación y las subbibliotecas.....	28
Tabla 6: Resultados de la validación interna. Se muestran los resultados promedio de las distancias para cada lote. Todos los valores de Hit son inferiores al umbral de cualificación 2.1 para la clase de Producto acabado (0,38487). ....	29
Tabla 7: Resultados de la validación externa. Especificidad y robustez. Se muestran los resultados promedio de las distancias para cada lote. Todos los valores de Hit son inferiores al umbral de cualificación 2.1 para la clase de Producto acabado (0,38487). ....	29
Tabla 8: Resultados de la validación externa con comprimidos varios. La identificación como producto acabado es negativa. Algunos productos se discriminan en el umbral de cualificación 1.1 (0,046877) y otros en la cualificación 2.1 (0.38487). ....	29
Tabla 9: Biblioteca espectral ampliada. Parámetros de la biblioteca. ....	30
Tabla 10: Valores de RMSEP y R2 en función del número de componentes PLS. ....	34
Tabla 11: Evaluación del error estándar de laboratorio para el método de referencia. ....	35
Tabla 12: Evaluación de la exactitud. Resultados de cuantificación de API por ambos métodos. ....	35
Tabla 13: Evaluación de la repetibilidad. Resultados del análisis de un sólo comprimido. ....	36
Tabla 14: Valores de Fexp y Ftab para los factores Día y Operador. ....	36
Tabla 15: Resultados en % de los análisis de una misma muestra en días diferentes por dos analistas distintos. ....	36
Tabla 16: Resultados obtenidos en la validación del modelo de calibración construido para cuantificar API en la especialidad farmacéutica de comprimidos recubiertos (expresados en %). ....	37
Tabla 17: Resultados de la comparativa de datos apareados para los equipos de producción y de laboratorio. ....	38
Tabla 18: Resultados obtenidos en la validación del modelo de calibración construido para cuantificar API en la especialidad farmacéutica en forma de granulado efervescente (expresados en %). ....	40

## Ilustraciones

Ilustración 1: Componentes de la radiación electromagnética.....	6
Ilustración 2: Espectro de la radiación electromagnética.....	6
Ilustración 3: Excitación de las vibraciones.....	7
Ilustración 4: Características principales de un espectrofotómetro NIR.....	8
Ilustración 5: Modo de registro en NIR.....	8
Ilustración 6: Generación de un haz de luz NIR mediante un interferómetro.....	8
Ilustración 7: Representación gráfica de la reducción de la dimensionalidad llevada a cabo por un PCA desde un sistema de 3 variables a uno de 2 componentes principales.[7].....	13
Ilustración 8: Diferentes situaciones en la diferenciación de las clases tras una validación interna.....	15
Ilustración 9: Función suma de dos tendencias: error de modelado y error de la estimación. ...	20
Ilustración 10: Espectros de todas las clases de la biblioteca en rango.....	28
Ilustración 11: Espectros de una muestra de producción de producto acabado y de principio activo en un rango de 12500-4000 cm <sup>-1</sup> . ....	32
Ilustración 12: Espectros del conjunto de calibración (en unidades de absorbancia).....	33
Ilustración 13: Espectros del conjunto de calibración (en unidades de absorbancia).....	33
Ilustración 14: Comparación de los espectros de las materias primas usadas en la fabricación de la especialidad farmacéutica en granulado efervescente. ....	39

## Gráficos

Gráfico 1: Gráfico de scores en primera derivada de muestras de laboratorio y muestras de producción. Los gráficos de scores en normalización vectorial o en segunda derivada dan información análoga. ....	31
Gráfico 2: Selección de los conjuntos de validación y calibración. Gráfico scores de los dos subconjuntos. Los espectros fueron sometido a una segunda derivada (9 puntos de suavizado) en un rango de 9200 a 4000 cm <sup>-1</sup> . ....	32
Gráfico 3: Gráfico RMSEP según el número de componentes PLS. ....	33
Gráfico 4: Gráfico de R <sup>2</sup> según el número de componentes PLS. ....	34
Gráfico 5: Root Mean Standard Error of Prediction vs núm de PCs (Izquierda) y Coeficiente de la regresión al cuadrado vs el núm de PCs. ....	39

## Diagramas:

Diagrama 1: Conjunto de biblioteca y subbibliotecas.....	27
Diagrama 2: Construcción de la biblioteca de identificación. ....	27

## RESUMEN

La industria farmacéutica ha mostrado especial interés en la técnica NIRS (espectroscopia de infrarrojo cercano) debido a su falta de preparación de muestra y a su carácter no destructivo y especialmente a su habilidad de llevar a cabo ambos análisis cualitativos y cuantitativos.

Se propone la construcción de bibliotecas espectrales del infrarrojo cercano como métodos alternativos para identificar primeras materias farmacéuticas. Se pueden identificar compuestos con espectros muy semejantes mediante la construcción de subbibliotecas derivadas de la principal que son desarrolladas mediante el uso de tratamientos quimiométricos con altas habilidades de discriminación.

Usar NIRS para análisis cuantitativos requiere de la aplicación de técnicas de calibración multivariable y de una selección cuidadosa y representativa de las muestras con el fin de construir modelos de calibración a partir de los espectros NIR usando la regresión parcial por mínimos cuadrados.

El objetivo de este trabajo es describir y construir un método para evaluar la identidad y el % de principio activo presente en un producto farmacéutico, y también conseguir la validez de un método ya construido (mediante el equipo NIR del laboratorio) en un equipo NIR nuevo de la planta de producción.

## ABSTRACT

The pharmaceutical industry has shown special interest in NIRS technique (near infrared spectroscopy) due to its lack of sample preparation and non-destructive character and especially to its ability to carry out both qualitative and quantitative analysis.

The construction of near-infrared spectral libraries as an alternative to qualitative analysis methods for identifying pharmaceuticals raw materials is proposed. Compounds with very similar spectra can be identified by constructing sub-cascading libraries derived from the main one that are developed by using chemometric treatments with high discriminating skills.

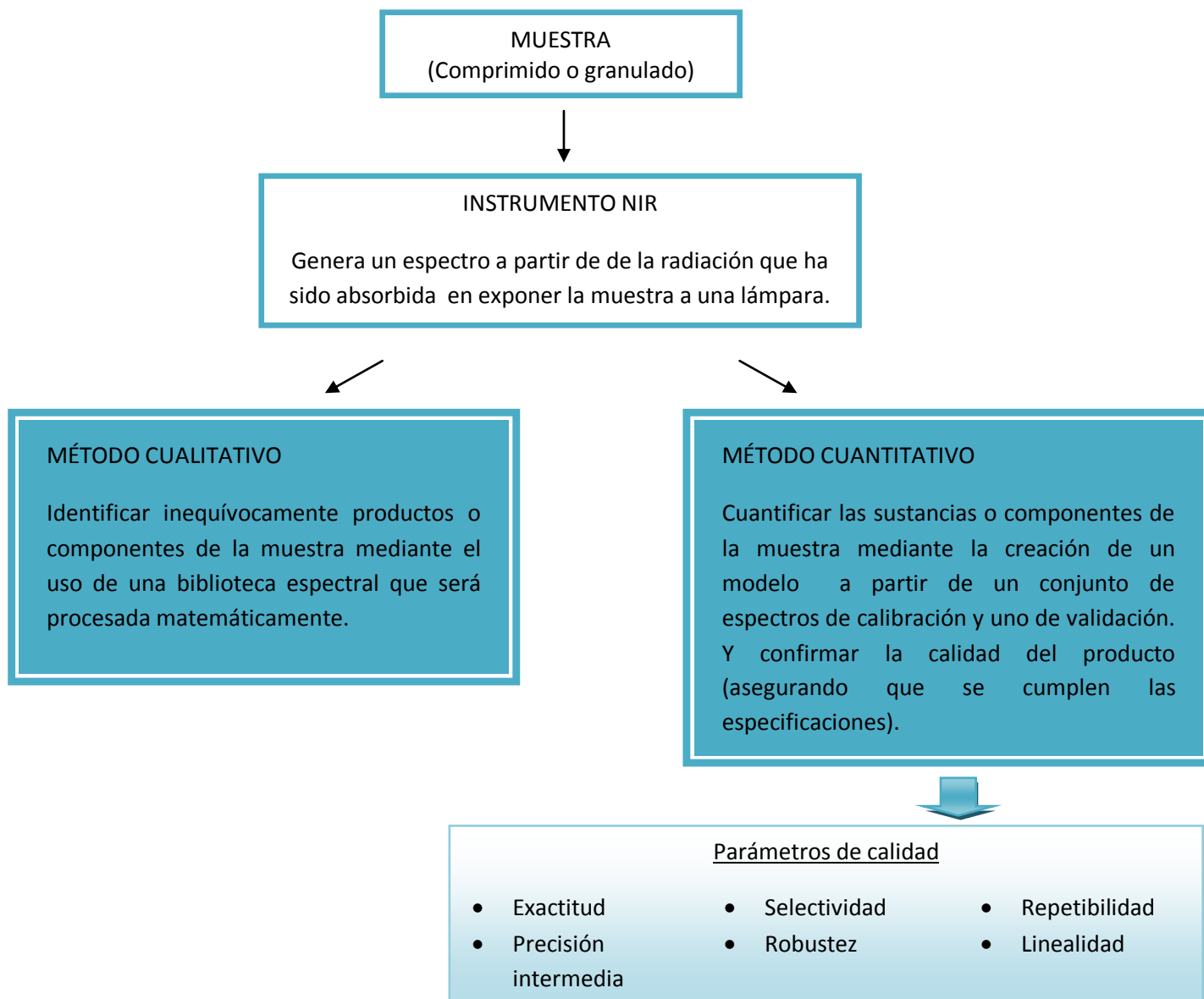
Using NIRS for quantitative analysis involves the application of multivariate-calibration techniques and careful selection of representative samples in order to construct calibration models from NIR spectra using Partial Least Squares Regression.

The aim of this work is to describe and build a method for asses the identity and the % of active pharmaceutical ingredient present in a pharmaceutical product, and also achieve the validity of one method already constructed (by means the NIR laboratory equipment) in one new NIR equipment in the production plant.

## 1. OBJETIVOS

El objetivo último de este trabajo es el desarrollo de métodos quimiométricos de identificación y cuantificación a partir de datos espectroscópicos en la región del infrarojo cercano. Más concretamente en este trabajo se pretende:

1. Describir la forma de construir y validar modelos de identificación y cuantificación para la determinación del contenido en % en API, mediante la espectroscopia NIR, en el producto farmacéutico en forma de comprimidos recubiertos



2. Validar la transferencia de calibración del método de identidad y cuantificación de una especialidad farmacéutica en forma de granulado efervescente en sobres entre los instrumentos NIR 21432 (Laboratorio) y 21479 (Planta Farmacéutica).

## 2. INTRODUCCIÓN A LOS MÉTODOS ESPECTROSCÓPICOS

Los métodos espectroscópicos son un amplio grupo de métodos analíticos que se basan en las espectroscopias atómica y molecular. La espectroscopia es el término general para la ciencia que trata de las distintas interacciones de la radiación con la materia.

La radiación electromagnética puede entenderse a partir de los parámetros de onda sinusoidal clásicos (como longitud de onda, frecuencia, amplitud de onda y velocidad), se representa como un campo eléctrico y otro magnético que están en fase, con oscilaciones sinusoidales en ángulo recto de uno respecto a otro y respecto

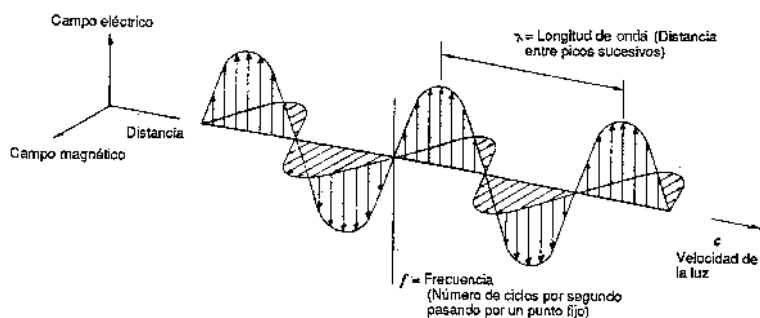


Ilustración 1: Componentes de la radiación electromagnética.

a la dirección de propagación; y no necesita un medio de apoyo para transmitirse. La componente eléctrica de la radiación es la responsable de la mayoría de los fenómenos que interesan como la transmisión, la reflexión, la refracción y la absorción. [1]

El espectro electromagnético abarca un intervalo enorme de ondas y de frecuencias y se divide en regiones espectrales determinadas según los métodos que se precisan para generar y detectar las diversas clases de radiación.

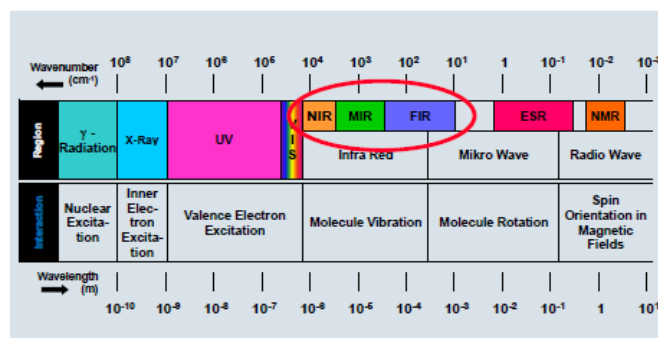


Ilustración 2: Espectro de la radiación electromagnética

### 2.1. Absorción de la radiación

Cuando la radiación atraviesa una capa de sólido, un líquido o un gas, ciertas frecuencias pueden ser selectivamente eliminadas por *absorción*, un proceso en el que la energía electromagnética se transfiere a los átomos, iones o moléculas que componen la muestra. La absorción provoca que estas partículas pasen de su estado normal a temperatura ambiente, o estado fundamental, a uno o más estados excitados de energía superior. Con este fin se realiza experimentalmente una representación gráfica de la absorbancia en función de la longitud de onda o frecuencia. La ley de Lambert-Beer expone que para una radiación monocromática, la absorbancia es directamente proporcional al camino óptico  $b$  a través del medio y la concentración  $c$  de la especie absorbente y  $\epsilon$  es la constante de absorptividad molar: [1]

$$A = \epsilon bc = -\log T = \log I_i/I_f$$

## 2.2. Requisitos para la absorción en la región del infrarrojo

- La radiación debe tener la energía precisa para satisfacer los requerimientos del paso al estado excitado (cuantización)
- Debe producirse una interacción efectiva entre la radiación y la materia (debe haber un cambio neto del momento dipolar como resultado de la vibración o rotación de la molécula para que se verifique la absorción)

## 2.3. Fundamento de la absorción en NIR

La vibración provoca fluctuaciones regulares del momento dipolar lo que origina un campo eléctrico que puede interactuar con el campo asociado a la radiación electromagnética. Cuando la frecuencia de la radiación es igual a la frecuencia de vibración natural de la molécula, se verifica una transferencia de energía que da lugar a un cambio de amplitud de la vibración molecular que lleva asociada una absorción de la radiación. [1]

Cuando se trata de especies homonucleares como el  $O_2$ ,  $N_2$  o  $Cl_2$ , el momento dipolar no sufre un cambio neto durante la vibración o la rotación y, como consecuencia, este tipo de compuestos no absorbe en el infrarrojo. Con la excepción de algunos compuestos de este tipo, todas las especies absorben radiación en el infrarrojo.

La región espectral del infrarrojo cercano (NIR) se extiende desde el extremo superior de longitudes de onda del visible, a alrededor de 770 nm, hasta 2500nm.

El origen de las bandas de absorción en el NIR son sobretonos (transiciones energéticas en las cuales la variación del número cuántico vibracional,  $n$ , es mayor de  $\pm 1$ ) y combinaciones lineales de las bandas vibracionales de tensión fundamentales (en moléculas poliatómicas y debidas a cambios simultáneos en la energía de uno o más modos de vibración), a diferencia de lo que ocurre en el infrarrojo medio e infrarrojo lejano donde la absorción se debe a vibraciones moleculares fundamentales y a rotaciones moleculares respectivamente. [7]

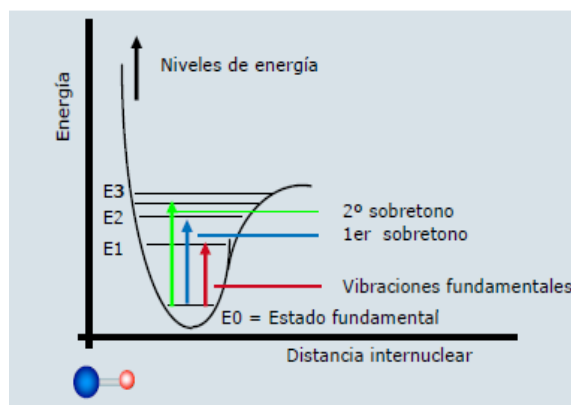


Ilustración 3: Excitación de las vibraciones.

## 2.4. Instrumentación para NIR

La baja intensidad de las bandas NIR, hace que el nivel de exigencia de los espectrómetros NIR, en términos de nivel de ruido permisible y estabilidad instrumental, deba ser mayor que en otros instrumentos, sobre todo si se pretende aplicar al análisis cuantitativo. El esquema básico de un espectrofotómetro NIR consta de: fuente de radiación, selector de longitudes de onda, el compartimento de la muestra y el detector. Las diferentes opciones para cada componente se muestran en la ilustración 4. [8]

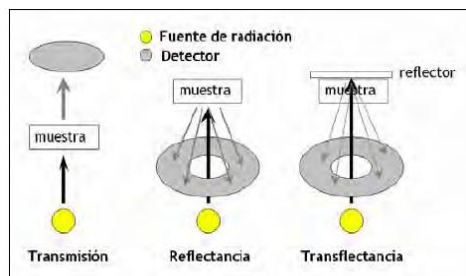


Ilustración 5: Modo de registro en NIR.

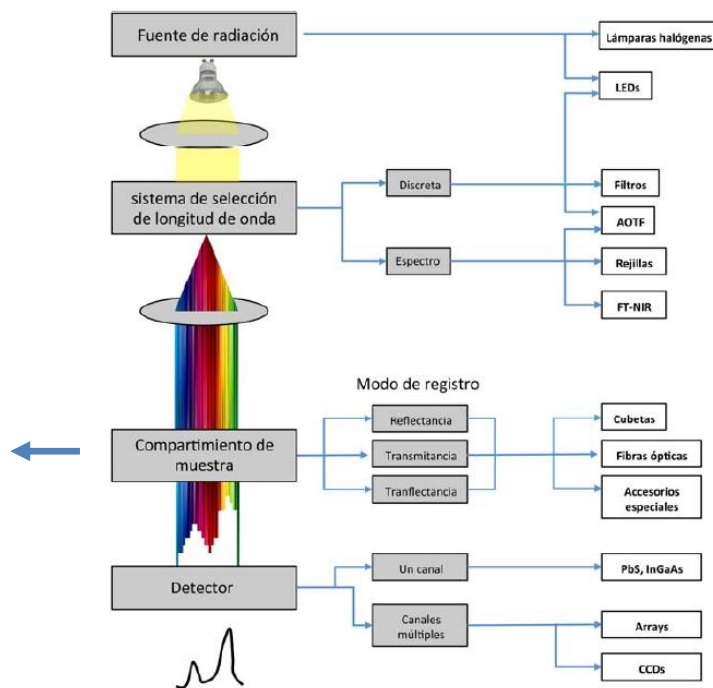


Ilustración 4: Características principales de un espectrofotómetro NIR.

### 2.5. Funcionamiento del FT-NIR de Bruker:

El principio de funcionamiento consiste en el envío de la radiación proveniente de la lámpara de tungsteno situada en el interior de la óptica del instrumento. Esta radiación llega al interferómetro que modula esta luz según el movimiento de unos espejos móviles. Esta luz modulada incide sobre la muestra saliendo de la óptica a través de una ventana de cuarzo y se detecta o bien la luz transmitida ( $I_t$ ) en el detector del módulo de transmisión o bien la luz reflejada en el detector del módulo de reflexión. En ambos casos se usa el mismo carrusel de muestra. La señal es un interferograma donde la intensidad se expresa en función del desplazamiento de los espejos móviles. A esta señal se le aplica automáticamente la transformada de Fourier que convierte esta intensidad en función de la frecuencia.

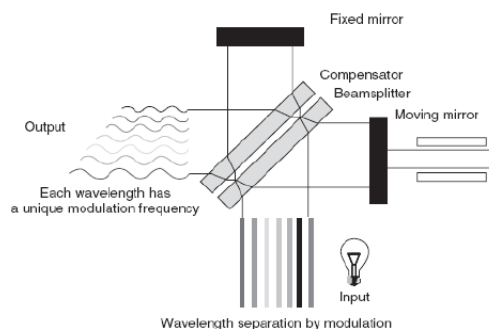


Ilustración 6: Generación de un haz de luz NIR mediante un interferómetro.

Para la obtención del espectro debe adquirirse primero la señal de referencia ( $I_i$ ) para eliminar la influencia de la propia óptica del equipo y efectos ambientales.

El resultado final es un espectro de transmitancia  $T = I_t/I_i$  o su equivalente de absorción  $A = -\log T$ . La adquisición de un conjunto de espectros de muestras de identidad conocida o bien de concentración conocida permite la creación de un modelo de identificación o de cuantificación mediante el software quimiométrico del instrumento.

### 3. QUIMIOMETRÍA.

La Quimiometría mejora el rendimiento del proceso analítico en todas sus etapas y asegura la calidad de los resultados. Las técnicas instrumentales, y especialmente las espectroscópicas, generan una gran cantidad de datos relativos a las muestras analizadas en breves intervalos de tiempo. La quimiometría pretende convertir esos datos en información mediante el uso de: [7]

- **Filtrado de señales** (pretratamientos aplicados a los datos experimentales)
- **Métodos de reconocimiento de pautas** (técnicas multivariantes mediante las que se pueden establecer agrupaciones de muestras en función de su similitud y métodos de clasificación de nuevas muestras).
- **Métodos de calibración** (técnicas multivariantes mediante las cuales se busca una relación cuantitativa entre la señal analítica y alguna propiedad de la muestra)

#### 3.1. Análisis Multivariable.

Tradicionalmente la manera de abordar un problema era analizarlo según una variable cada vez. Para algunas situaciones este enfoque es perfectamente aceptable, especialmente cuando las variables se relacionan una a una. Sin embargo, cuando estas relaciones son más complejas, una única variable no puede describir adecuadamente el sistema. Aquí es donde el Análisis multivariable es más útil.

El Análisis Multivariable es la investigación de muchas variables, simultáneamente, con el fin de entender las relaciones que existen entre ellas. Se trata de un enfoque muy gráfico ya que profundiza en la búsqueda de la información que está escondida detrás de los números.

El Análisis Multivariable es el estudio de la variabilidad y sus fuentes. La variabilidad puede ser definida como deseada y no-deseada:

- *La variabilidad deseada* es el cambio premeditado de parámetros físicos o químicos que afectan al sistema de estudio con el fin de evaluar el impacto que dichos parámetros ejercen sobre él
- *La variabilidad no deseada* resulta de la incapacidad para controlar algunos aspectos del sistema y es aleatoria.

Un modelo es un resumen del conocimiento que se tiene sobre un sistema en el momento de investigarlo. Los modelos se utilizan para predecir eventos futuros con una cierta confianza. Validar un modelo implica demostrar que se adecua al uso para el que fue pensado.

Un modelo multivariable es capaz de mostrar la influencia que tiene en el sistema los dos tipos de variabilidad (deseada y no deseada) de modo que éste puede ser mejor comprendido. [5]



Beneficios del Análisis Multivariable (MVA): [5]

- MVA identifica variables que contribuyen a la mayoría de la variabilidad de los datos.
- MVA ayuda a aislar aquellas variables que están relacionadas, en otras palabras, que co-varían las unas con las otras.
- La manera gráfica de presentar los resultados permite una mejor interpretación.

En este trabajo el uso del análisis multivariable ha sido indispensable para la creación de modelos de identificación y cuantificación con el objeto de desarrollar métodos analíticos para determinar el contenido de principio activo en las especialidades farmacéuticas.

3.1.1. Método cualitativo: creación de una biblioteca espectral

La información química se extrae de los espectros NIR de materias primas, del producto semielaborado y del producto acabado. Un espectro de una sustancia puede ser comprendido como un vector, es decir, una matriz de una fila (porque hace referencia al espectro de una sola muestra) con m columnas (tantas columnas como longitudes de ondas se irradian en la muestra). Para cada longitud de onda, es decir, cada columna, corresponde un valor de absorbancia.

La biblioteca espectral que se desea crear debe contener información de todos los excipientes del producto, del principio activo (API) y del producto acabado en diferentes fases de su producción. En otras palabras, la biblioteca espectral debe recoger toda la variabilidad del sistema: tanto la química (registro de diversas sustancias y mezclas) como la física (mezcla en polvo, comprimidos recubiertos y comprimidos sin recubrir).

Toda esa información se traduce en una matriz, que proviene de todos los espectros registrados, con n filas (tantas como muestras registradas) y m columnas, tantas como longitudes de onda hayan sido utilizadas.

Es aquí donde el análisis multivariable surge como una poderosa herramienta para obtener información de todo ese cúmulo de datos, ya que podemos estar contemplando miles de variables y un análisis convencional univariable no sería exitoso.

Previo a este análisis es conveniente procesar los espectros obtenidos con algunos pretratamientos matemáticos que permitan aumentar la relación señal/ruido y por ende acrecentar la señal correspondiente al analito de interés.

La elección de un tratamiento matemático u otro depende de las características de las muestras a clasificar, se escoge aquel tratamiento matemático que proporcione mejor discriminación entre las muestras, en otras palabras, la selección del pretratamiento más apropiado es experimental (a base de ensayo y error).

Los pretratamientos espectrales más utilizados son: [6], [7]

### **PROMEDIO DE ESPECTROS**

Cada medida realiza un nº alto de registros (el promedio minimiza el ruido) → finalmente se obtiene un espectro.

### **CENTRADO**

Consiste en restar a cada valor de cada variable el valor medio de la misma. Así el valor de cada variable se encuentra referido al centro del modelo:

$$x' = x_{i,j} - \bar{x}_j$$

Donde,  $x'$  es el dato centrado,  $x_{i,j}$  es el dato del espectro  $i$  y la variable  $j$  antes del centrado y  $\bar{x}_j$  es el promedio de la columna  $j$

### **AUTOESCALADO**

Una vez centrados los datos se dividen por la desviación estándar de la variable correspondiente.

Todas las variables tienen media = 0 y desviación estándar = 1

$$x' = \frac{x_{i,j} - \bar{x}_j}{s_j}$$

Donde  $x'$  es el autoescalado,  $x_{i,j}$  es el dato de la fila  $i$  (espectro  $i$ ) y la columna  $j$  (variable  $j$ ) antes del autoescalado,  $\bar{x}_j$  es el promedio de la columna  $j$  y  $s_j$  es la desviación estándar de la columna.

Nota: el software quimiométrico asociado al instrumento NIR utilizado realiza el centrado y autoescalado de los datos por defecto.

### **NORMALIZACIÓN VECTORIAL (SNV, Standard Normal Variate)**

Se utiliza para corregir el efecto de dispersión producido por ejemplo, por las diferencias espectrales en el tamaño de partícula de muestras sólidas ya que minimiza variaciones de desplazamiento de la línea base. Se basa en restar cada variable obtenida, absorbancia, del promedio y dividir el valor entre la desviación estándar.

$$X_{i,j}^{SNV} = \frac{x_{i,j} - \bar{x}_i}{s_i}$$

Donde  $x'$  es el autoescalado,  $x_{i,j}$  es el dato de la fila  $i$  (espectro  $i$ ) y la columna  $j$  (variable  $j$ ) antes del autoescalado,  $\bar{x}_i$  es el promedio de la fila  $i$  y  $s_i$  es la desviación estándar de la fila.

## DERIVADAS

El uso de derivadas permite aumentar las diferencias entre las bandas anchas y solapadas del espectro. La primera derivada minimiza diferencias de desplazamiento de la línea base y la segunda derivada minimiza diferencias de pendiente. La elección de más o menos puntos de suavizado tendrá que evaluarse según cada caso, ya que ventanas grandes hacen que pequeñas diferencias en los espectros queden diluidas, aunque también hacen que el ruido sea menor.

Después de procesar las señales, se pueden aplicar dos métodos multivariantes que ayudan a clasificar los datos en grupos o patrones y a identificar variables que influyen en el sistema, gracias a que permite la visualización de tendencias en los datos: Análisis de Clusters y Análisis en componentes principales (PCA). En la creación de la biblioteca espectral se ha usado ampliamente el segundo.

### 3.2. Análisis de Clusters (o conglomerados) [7][2]

Es un método para dividir un grupo de objetos en una serie de clases de manera que los objetos similares se encuentren en la misma clase. El análisis de clusters busca objetos que se encuentren próximos en el espacio de las variables. La distancia,  $d$ , entre dos puntos en un espacio  $n$ -dimensional con coordenadas  $(x_1, x_2, \dots, x_n)$  y  $(y_1, y_2, \dots, y_n)$  se calcula habitualmente a través de la **distancia Euclídea** definida por:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

### 3.3. Análisis en componentes principales (PCA) [4][11][12]

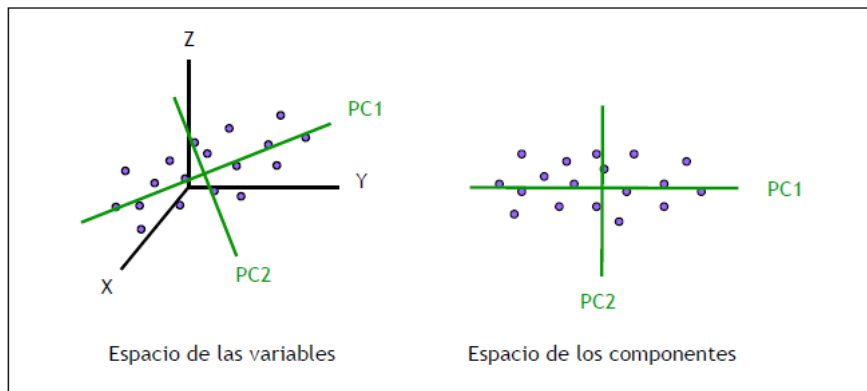
PCA es una técnica multivariable en la cual un número de variables más o menos correlacionadas son transformadas a otro número menor de variables no correlacionadas. Esta colinealidad significa que la matriz de datos  $X$  tiene algún tipo de variabilidad dominante que lleva la información principal; el objetivo de PCA es eliminar la información redundante y la variabilidad debida al ruido.

El punto de partida en todos los análisis multivariable es una matriz de datos  $X$ . En espectroscopia, esa matriz puede corresponder a los espectros de un conjunto de muestras; cada objeto (fila de la matriz) corresponde al espectro de una muestra, y las columnas son las longitudes de onda a las que ha sido registrado el espectro. Si registramos  $M$  muestras a  $K$  longitudes de onda tendremos una matriz de dimensión  $M \times K$ .

Un espectro, medido a  $K$  longitudes de onda, constituye un conjunto de  $K$  variables que puede ser descrito matemáticamente como un vector. Se puede construir un espacio de  $K$  dimensiones de forma que cada dimensión corresponda a una longitud de onda, y se puede representar el mismo objeto como un punto en este nuevo espacio. Si tenemos  $M$  muestras, se pueden representar como otros tanto puntos en el espacio de  $K$  dimensiones. Si no tienen nada en común los unos con los otros, los  $M$  puntos estarán dispersos en el espacio, pero si los  $M$  espectros están relacionados (por ejemplo espectros de mezclas en dos componentes en diferentes proporciones) los  $M$  puntos aparecerán agrupados.

El objetivo de PCA es hallar las direcciones en que están agrupados los  $M$  puntos de los objetos en el espacio de  $K$  variables, y reducir el sistema inicial  $K$  dimensional a uno  $A$  dimensional, con  $A < K$ , manteniendo la información relevante del sistema. Geométricamente es un cambio de ejes, representando esas muestras en un nuevo sistema de coordenadas con un número inferior de ejes al utilizado inicialmente.

El análisis en componentes principales busca las direcciones que expliquen la máxima variabilidad de las muestras, definiendo unos nuevos ejes que son los componentes principales. El primer componente principal es la dirección que explica la máxima variabilidad de las muestras; el segundo, se escoge que sea perpendicular al primero, y explique la máxima variabilidad de las muestras una vez restada la explicada por el primer componente principal, y así, sucesivamente.



**Ilustración 7:** Representación gráfica de la reducción de la dimensionalidad llevada a cabo por un PCA desde un sistema de 3 variables a uno de 2 componentes principales.[7]

Para poder definir esos ejes matemáticamente se utilizan los cosenos de los ángulos que forman cada uno de estos nuevos ejes con los antiguos, son los loadings, y las coordenadas de las muestras en estos nuevos ejes son los scores.

Matemáticamente la matriz de datos  $X$  se descompone en el producto de dos matrices,  $T$  (matriz de scores) y  $P$  (matriz de loadings), más una matriz  $E$  de residuales de  $X$ :

$$X = TP^T + E$$

$$\begin{array}{c}
 \boxed{X} = \begin{array}{c} \boxed{t_1} \\ \vdots \\ \boxed{t_p} \end{array} \begin{array}{c} \boxed{\begin{matrix} 1 & \dots & p \\ p_1 & & \end{matrix}} \\ \vdots \\ \boxed{\begin{matrix} 1 & \dots & p \\ p_2 & & \end{matrix}} \\ \vdots \\ \boxed{\begin{matrix} 1 & \dots & p \\ \vdots & & \end{matrix}} \end{array} + \dots + \boxed{E}
 \end{array}$$

Los nuevos ejes son los componentes principales, las nuevas coordenadas son los scores y la relación entre ejes antiguos y nuevos ejes son los loadings. Los scores contienen información sobre los objetos (muestras) y los loadings contienen información sobre las variables.

Algunas guías generales para la aplicación de PCA: [5]

- Siempre examinar en primer lugar el gráfico de la varianza explicada antes de ver cuántos PCs son requeridos para describir una porción razonable de variabilidad. Cuántos menos se necesiten, mejor explicado estará el sistema. Si se toman muchos PCs para describir menos del 50% de los datos, el conjunto de datos contiene básicamente ruido.
- Después de decidir cuántos PCs deben ser usados (del gráfico de la varianza explicada), examinar el gráfico de scores representando PC1 vs PC2 (y otras combinaciones de todos los PCs que hayan decidido usarse) y mirar tendencias. Estos patrones pueden revelar pequeñas estructuras de datos relacionadas con otros fenómenos que ocurren en el sistema.
- Usar el gráfico de loadings para interpretar qué variables contribuyen a la mayoría de clases de muestras (si existe alguna). Esto puede ayudar a tener una mejor percepción de lo que ocurre en el sistema y también revela si hay alguna variable que interaccione con el resto. (Importante en el control de procesos, el gráfico de loadings corresponderá al de la variable que cambia durante el proceso)
- Usar el gráfico de influencia para determinar si las muestras están bien descritas por el modelo usando leverage values para concluir si hay muestras extremas presentes. (Las muestras extremas tienen un gran peso en el modelo, representado el gráfico de influencia, residuales vs leverage values, podremos descartar muestras extremas con residuales elevados).

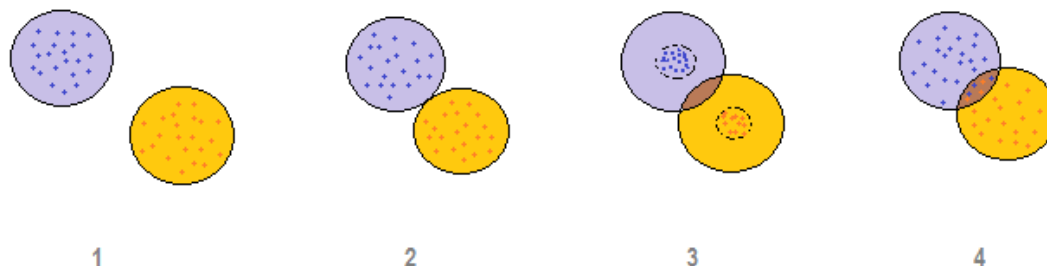
El uso de pretratamientos espectrales y un método de reducción de variables en el desarrollo de la biblioteca espectral, conduce a la diferenciación de todas las clases que componen la matriz inicial de los datos. Se establece para cada clase un valor umbral:

$$\text{Umbral} = \text{Hit máximo} + X \cdot SD$$

siendo *Hit máximo* la distancia euclídea entre el espectro promedio de la clase y el espectro más alejado perteneciente a la clase. *X* es un factor multiplicativo, el valor por defecto es 0.25 y es equivalente a asignar un nivel de confianza del 99.99% y *SD* es la desviación estándar del conjunto de espectros que conforman la clase.

Idealmente, cuando la biblioteca es validada internamente (mediante las muestras usadas en la construcción de la biblioteca), todos los espectros se distinguen bien de otros de diferentes clases y los umbrales son tan pequeños que sus regiones de confianza no se solapan. Los espectros son puntos (en un sentido matemático) en un espacio dimensional *n*, siendo *n* el número de puntos de datos seleccionado. Las regiones de confianza se pueden representar por esferas, el centro de las cuales representa el espectro promedio de cada clase. [6]

Sin embargo, puede ocurrir que las regiones de confianza de algunos espectros promedio se solapen o simplemente se confundan. Se distinguen 4 situaciones a las que el software quimiométrico del equipo NIR otorga 4 colores distintos en el resultado de la validación interna de la biblioteca.



**Ilustración 8: Diferentes situaciones en la diferenciación de las clases tras una validación interna.**

En el primer caso, los espectros se diferencian bien, los resultados de la validación interna salen en color verde.

En el segundo caso, no hay ningún espectro de los que se ha registrado para una clase que esté dentro de la región de confianza de otra, pero éstas se confunden. Se otorga un color negro a los resultados de la validación interna.

En el tercer caso, las regiones de confianza se solapan, aunque no hay ningún espectro (de ninguna de las dos clases) que se encuentre en esa región compartida. Los resultados de la validación interna poseen un color gris.

En el último caso, las regiones de confianza se solapan de manera tal que las clases no se pueden diferenciar; los resultados en este caso son rojos.

La biblioteca espectral desarrollada con estas herramientas deberá ser validada externamente (mediante otras muestras de producto acabado y comprimidos/granulados de otras especialidades farmacéuticas) con tal de constatar que el método de identificación desarrollado se ajusta al objetivo para el cual fue pensado.

### 3.3.1 Método cuantitativo: Uso de Análisis de Regresión y modelos predictivos

La calibración se define clásicamente en Química Analítica como el conjunto de operaciones que establecen, en las condiciones específicas, la relación entre los valores indicados por un instrumento de medida o por un sistema de medida y los valores conocidos correspondientes a un valor medido. [4]

La fórmula matemática que relaciona las variables de respuesta o señales analíticas con las concentraciones es la ecuación de calibración o modelo.

El número de variables utilizadas es un aspecto crítico de todas las alternativas matemáticas que relacionan la respuesta del instrumento con el parámetro que se desea determinar.

El objetivo de la calibración multivariable es la determinación simultánea de varios analitos o la de un único analito en presencia de una matriz compleja que participa en la señal analítica.

El Análisis de Regresión es el proceso de desarrollar un modelo a partir de los datos disponibles para predecir una respuesta deseada (o respuestas). El análisis de regresión requiere dos tablas de datos, una de variables independientes y otra de variables dependientes. [5]

- Variables independientes** son el conjunto de medidas disponibles que se usan para que un modelo prediga la respuesta de interés.
- Variables dependientes** son las respuestas que se intentan extraer del modelo de predicción construido a partir de las variables independientes.

### 3.4. Métodos de regresión multivariable [10][9][6][3][5][2]

Hay tres tipos de regresión multivariable comúnmente usados:

- Regresión lineal múltiple (MLR)**
- Regresión en componentes principales (PCR)**
- Regresión parcial por mínimos cuadrados (PLSR)**

Cuando los espectros (es decir, variables fáciles de medir) son pocos en número, no son significativamente redundantes (colineales) y tienen una relación bien conocida con la respuesta (concentración), entonces la regresión lineal múltiple (MLR) puede ser una buena manera de convertir esos datos en información. Sin embargo si una de esas tres condiciones se rompe, MLR puede ser ineficiente o inapropiada ya que probablemente se obtendrá un modelo que define perfectamente las muestras pero que fallará en la predicción de datos nuevos (sobreajuste = over-fitting). [9]

Puede darse el caso en el que el investigador se enfrente con muchas variables o con muchas relaciones pobremente comprendidas. En esas situaciones son útiles los dos últimos métodos, que son conocidos como métodos de regresión latentes, los cuales tienen la ventaja de revelar la estructura escondida en el conjunto de datos. Basados en un método de reducción de variables, estos métodos implican una reducción de la dimensionalidad y una simplificación de la interpretación química.[9]

Como la PCR, la regresión parcial sobre mínimos cuadrados (PLSR) empieza encontrando combinaciones lineales de las variables predictoras (factores o componentes principales). Sin embargo, la forma en que se eligen esas combinaciones resulta diferente. En PCR los componentes principales se eligen de manera que describan tanto como sea posible la variación entre las variables predictoras, al margen de la fuerza de las relaciones entre las variables respuesta y las predictoras. En PLSR, a las variables que muestran una alta correlación con las variables respuesta se les asigna un peso extra dado porque serán más efectivas en la predicción. De esta forma se eligen combinaciones lineales de las variables predictoras que estén altamente correlacionadas con las variables respuesta y también expliquen la variación en las variables predictoras. [2]

En otras palabras, MLR podría asemejarse a un conjunto de  $n$  ecuaciones con  $n$  incógnitas, en nuestro caso, parecido a la aplicación de Lambert-Beer para mezclas, y válido siempre y cuando ese conjunto de ecuaciones sea discreto. PCR es capaz de analizar conjuntos de datos significativamente más extensos y se basa en un análisis de componentes principales, buscando la máxima variabilidad entre los espectros. En cambio, PLSR aunque también es capaz de analizar muchas variables, el método de reducción usado es distinto en que se busca la máxima variabilidad asociada a un cambio en la concentración (que puede no ser la máxima variabilidad del sistema).

Los modelos de cuantificación desarrollados en este trabajo constan de una primera etapa de calibración (desarrollo del modelo), una de verificación (validación del modelo) y una última de predicción. Hay que tener en cuenta que el objetivo no es encontrar el mejor modelo de calibración sino el modelo con la mejor capacidad predictiva.

### 3.5. Validación de métodos analíticos

Validar un método de análisis consiste en verificar y documentar su validez, esto es, su adecuación a unos determinados requisitos establecidos que determinan, a su vez, los criterios de calidad del método a utilizar. Los **criterios de calidad** pueden ser de tipo estadístico (exactitud y precisión en primera instancia y selectividad, sensibilidad, límites de método y robustez en segundo lugar) o de tipo económico/operativo (comprensión del método, manejo de la instrumentación, rapidez del análisis, coste y mantenimiento).[16]

Los principios comunes de la validación en análisis multivariable son:[5]

**Validación cruzada:** se define un conjunto de muestras que se deja de lado en la construcción del modelo, es decir, el modelo, se desarrolla con las muestras restantes. Las muestras que no se introdujeron en la construcción son usadas en la comprobación de la calidad del modelo. Este proceso se repite  $n$  veces de manera iterativa hasta que todas las muestras son usadas para la construcción y validación del modelo.

**Examen por conjunto de validación:** este método es la mejor manera de validar un modelo multivariable debido a que usa conjuntos separados de datos para construir y validar el modelo.

Los beneficios de la validación son:[5]

- Encontrar el modelo más simple y de más confianza
- Aislamiento de muestras que tienen una alta influencia en el modelo
- Mejor interpretación del modelo
- Tener la garantía de que la mayor variabilidad ha sido cubierta en el modelo, con respecto al conjunto de prueba.



## 4. METODOLOGÍA

### 4.1. Descripción de la creación del modelo de identificación

1. Registrar los espectros NIR usando un conjunto de muestras de identidad conocida.

2. Elección de los espectros.

Para cada componente o sustancia del producto farmacéutico, los espectros escogidos que conformen el conjunto de calibración deben representar toda la variabilidad química y física del sistema.

3. Construcción de la biblioteca.

Se deben escoger los parámetros de construcción de la biblioteca, tales como el pre-tratamiento espectral (SNV, derivadas...), el rango de longitudes de onda y el umbral.

Luego debe escogerse un método de reconocimiento de pautas (coeficiente de correlación, análisis de Clusters, PCA..)

El siguiente paso es una validación interna de la biblioteca, con tal de comprobar que no hay espectros mal identificados o que llevan a errores de identificación o ambigüedades. Basándose en los resultados de la validación interna, se pueden modificar los parámetros usados en la construcción de la misma a fin de obtener la que discrimine todas las sustancias con el grado de especificidad deseado.

4. Construcción de subbibliotecas.

Cada subbiblioteca debe incluir todas aquellas sustancias que sean similares debido a la semejanza de sus espectros. Los pasos necesarios para construir una subbiblioteca son los mismos que los descritos para construir una biblioteca y llevar a término los cálculos, variando los parámetros, a fin de que todos los componentes puedan ser inequívocamente identificados.

5. Validación externa.

El conjunto de toda la biblioteca (la principal y las subbibliotecas) debe ser externamente validada con muestras no incluidas en el conjunto de calibración que proporcionen resultados de identificación sin ambigüedades.

## 4.2. Descripción de la creación del modelo de cuantificación

### 1. Diseño de un conjunto de calibración.

Construir un conjunto de calibración adecuado requiere:

- a) Disponer de muestras con un rango suficientemente amplio de concentraciones con el fin de asegurar la capacidad predictiva del proceso
- b) Contemplar toda la variabilidad del proceso, la química y la física.

Para ello, en el laboratorio se preparan, por pesada, 30 mezclas en polvo de los excipientes y el API en un rango de concentración de 70-130% nominal del contenido en API. Para incluir la variabilidad física en el modelo de cuantificación, parte de las muestras del conjunto de calibración son muestras de producción del producto acabado. Además también se lleva a cabo una evaluación de la presión de compactación de los comprimidos de producción con tal de comprimir, en el laboratorio, muestras preparadas por pesada y con la concentración nominal de API del 100% en un rango de presiones similar al que es comprimido el producto acabado.

En conclusión, el conjunto de calibración lo conforman muestras de laboratorio comprimidas en un determinado rango de presiones y muestras de producción. Se conoce la concentración de las muestras de laboratorio por la pesada de los excipientes y el API, y la concentración de las muestras de producción se evalúa por el método de referencia establecido: HPLC (High Pressure Liquid Chromatography)

### 2. Registro de los espectros y aplicación del modelo quimiométrico de calibración

Se registra el espectro NIR de todas las muestras que conforman el conjunto de calibración. Los espectros deben registrarse con el mismo equipo que va a utilizarse en la predicción, después se procesan matemáticamente para evitar desplazamientos de la línea base o cambios en la pendiente y poder aumentar la relación señal/ruido. El procedimiento para seleccionar las muestras que forman los conjuntos de calibración y de validación se efectúa de modo que ambos subconjuntos contengan la variabilidad espectral y el rango de concentraciones antes indicado. Para seleccionar las muestras adecuadamente se realiza un análisis en componentes principales (PCA) de todo el conjunto de muestras. Los gráficos de scores muestran esa selección en función de la variabilidad.

Luego se aplica el algoritmo de regresión multivariable PLS y el rango espectral usado para la construcción de ese modelo PLS se selecciona basándose en las posiciones de las bandas de absorción del espectro de API puro, de esta manera se pueden excluir aquellas bandas del espectro NIR que son demasiado débiles o que únicamente proporcionan ruido espectral. Se deben detectar y descartar out-liers y verificar que el modelo mejora su capacidad predictiva. Aunque no debe descartarse tomar muestras en situaciones “no conformes” (es decir fuera de control estadístico) porque eso puede alimentar el modelo con una variabilidad importante para ampliar el rango de calibrado.

### 3. Validación del modelo.

La validación del modelo es el estudio cuantitativo de los resultados de aplicar el modelo construido con los datos de calibración a nuevas muestras. La validación llevada a término no ha sido una *cross-validation* sino mediante un conjunto de prueba. Una vez definidos los conjuntos de calibración y de prueba, el rango o rangos de longitudes de onda, el pretratamiento matemático y cálculo PLS con un máximo número de componentes, el software quimiométrico proporciona los parámetros de calibración y validación a la vez, pudiendo decidir entonces la dimensionalidad del modelo y la detección de out-liers mediante gráficos RMSEP vs Núm. de PCs,  $R^2$  vs Núm. de PCs, diferencia vs verdadero, predicción verdadero, score1 vs score 2...

### 4. Dimensionalidad del modelo.

La dimensionalidad del modelo viene determinada por el número de factores o componentes principales PLS. Un valor muy pequeño resultará en un modelo sin información suficiente para predecir adecuadamente los valores (under-fitting). Un valor demasiado alto hará que el modelo tenga demasiada información particular de algunas muestras (ruido o detalles menores) y, por tanto, la predicción de muestras desconocidas será defectuosa (over-fitting).

El número óptimo de factores que se debe escoger es aquel que proporcione un menor RMSE (Root mean Square Error of Prediction):

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^m (Y_i^{\text{NIR}} - Y_i^{\text{REF}})^2}{m}}$$

donde  $Y_i^{\text{NIR}}$  es el contenido en API predicho por el modelo NIR,  $Y_i^{\text{REF}}$  es el valor de referencia (obtenido por pesada y por HPLC) y  $m$  es el número de muestras. Se pretende encontrar un valor de RMSEP por debajo de 2 y está expresado con las mismas unidades que los datos de referencia.

Otros parámetros de predicción a tener en cuenta son: [4]

- **Bias:** desviación promedio sistemática entre los datos verdaderos y los predichos. Se calcula promediando las desviaciones particulares de las muestras dentro del conjunto de datos y se pretende que sea lo más parecida a cero.

$$\text{Bias} = \frac{\sum_{i=1}^m (Y_i^{\text{NIR}} - Y_i^{\text{REF}})}{m}$$

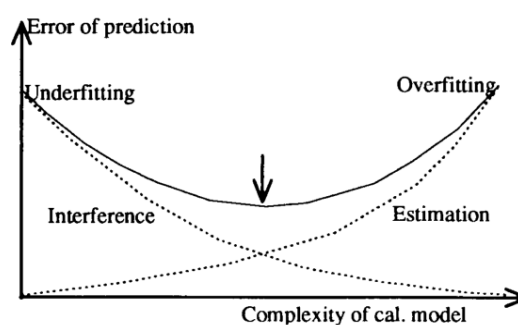


Ilustración 9: Función suma de dos tendencias: error de modelado y error de la estimación.

- **Residual Prediction Deviation (RPD):** es el cociente entre la desviación estándar de los valores de referencia (SD) y  $SEP_{bias}$  es el error promedio corregido con el bias de la predicción de la validación. Se busca un valor de RPD mayor de 5.

$$RPD = \frac{SD}{SEP_{bias}}$$

$$\left\{ \begin{array}{l} SD = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (Y_i^{NIR} - \bar{Y})^2} \\ SEP_{bias} = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (Y_i^{NIR} - Y_i^{REF} - bias)^2} \end{array} \right.$$

- **$R^2$ , o coeficiente de correlación**, proporciona el porcentaje de varianza presente en los valores de los componentes, que después son reproducidos en la predicción. También es denominado varianza explicada. Cuanto más alto sea el coeficiente, mejor será la correlación entre los datos de concentración y los datos espectrales.

#### 5. Parámetros estadísticos del modelo desarrollado.

A base de ensayo y error y mediante un minucioso trabajo de examen de tendencias, buena selección de rangos y de pretratamientos adecuados y descarte de out-liers se llega al modelo de cuantificación que reúne unos mejores parámetros de predicción (RMSEP, bias,  $R^2$ , número de componentes...)

Después de establecer el mejor modelo PLS de entre todos los creados, se deben definir unos parámetros de calidad del método analítico construido mediante NIR que lo hagan válido respecto cualquier otro método de referencia usado para cuantificar el contenido de API de los productos farmacéuticos observados.

Parámetro	Linealidad	Exactitud	Repetibilidad	Precisión intermedia	Solidez
<b>Procedimiento</b>	[NIR]= a + b [HPLC]  Se define una incertidumbre para a y para b ( que deben incluir el 0 y el 1 respectivamente)	Test t de datos emparejados de valores de NIR y valores de REF de los lotes de producción.	Las muestras de producción son analizadas seis veces por el mismo operador	Las muestras de producción se analizan 3 días por dos operadores diferentes. (ANOVA de dos factores)	Test t de datos emparejados entre valores de NIR y valores de REF de muestras de producción a lo largo del tiempo.

Tabla 1: Procedimientos para demostrar los parámetros de calidad del modelo, según la Agencia Europea del Medicamento (EMA). [15],[17]

## 5. PROCEDIMIENTO EXPERIMENTAL

### 5.1. Instrumentación

#### Espectrómetro

BRUKER MPA T nº de espectrómetro: 122000, equipado con tecnología de transformada de Fourier con esfera integradora y un automuestreador.

**Localización:** todas las medidas NIR fueron tomadas en las instalaciones de los Laboratorios Kern Pharma, S.L. en el Polígono Industrial Colon II, C/ Venus, 72, 08228 Terrassa (Barcelona).

**Fuente de emisión:** lámpara halógena.

**Interferómetro:** interferómetro patentado de transformada de Fourier RockSolid™, con alineación permanente, altamente estable e insensible a las vibraciones. Opera con un rango espectral de 12500-3750  $\text{cm}^{-1}$ , con una resolución espectral de pico de 2  $\text{cm}^{-1}$  y  $\pm 0.3\text{nm}$  de exactitud.

**Detector:** detector altamente sensible de PbS.

**Técnica de medida:** los datos espectrales son adquiridos desde 12,500 – 3750  $\text{cm}^{-1}$  con intervalos de 8  $\text{cm}^{-1}$ . Las reflexiones provenientes de la muestra son concentradas en la esfera integradora y finalmente alcanzan el detector, localizado en un lado de la misma. Este sistema de medida garantiza que ninguna de las reflexiones de la muestra se pierde, es decir, que no se pierde información.

**Software:** OPUS 7.0 y 7.2, proporcionado por Bruker Optics.

**Preparación de la muestra:** la muestra no necesita previa preparación ni derivatización. El análisis por esta técnica no es destructivo, así que la muestra puede ser seguidamente analizada por otras técnicas.

**Registro del background:** se toma previo a cualquier secuencia de medida y es válido durante 1 hora.

**Espectro de referencia:** la referencia de reflectancia es medida sobre una cobertura de oro integrada en el equipo.

**Registro del espectro de las muestras:** cada espectro se registra por duplicado y el promedio es usado en la predicción de la concentración. Los dos espectros adquiridos son un promedio, cada uno, de 32 scans. Las medidas NIR son adquiridas en 30 segundos.



Imagen 1: Configuración del equipo NIR. La fotografía muestra el espectrómetro y el automuestreador para comprimidos.

#### Equipo HPLC

Se lleva a cabo un análisis del contenido en principio activo mediante un equipo de HPLC Agilent (High Pressure Liquid Chromatography) acoplado a un detector ultravioleta. Los cromatogramas y la posterior cuantificación de principio activo fueron realizados en los laboratorios de KERNPHARMA S.L. (Terrassa, España).

## 5.2. Registro de espectros para el modelo de identificación.

La información química se extrae de los espectros NIR. Todas las muestras deberán registrarse en el mismo equipo siguiendo el mismo protocolo de análisis. Previo al registro de cualquier muestra es necesario tomar un espectro de referencia de la placa cerámica. Cada espectro de cada sustancia es dividido entre su espectro de referencia; esto permite eliminar cambios debidos, por ejemplo, al envejecimiento de la lámpara.

Para construir la biblioteca espectral se registran todas las materias primas que conforman la especialidad farmacéutica: registro de los excipientes y del API, así como también de la mezcla placebo (mezcla de todos los excipientes a excepción del API). Se registran los espectros de los productos intermedios (mezclas finales, núcleos en la fabricación de comprimidos y producto acabado).

Las materias primas (excipientes y principio activo), así como el placebo y la mezcla final (mezcla de todos los ingredientes después de ser granulada y tamizada y, en el caso de comprimidos, antes de ser prensada y recubierta) se analizan por triplicado mediante el instrumento NIR acoplado a un carrusel rotatorio en el que se incorporan las muestras en viales.

Los núcleos y el producto acabado en lugar de ponerse en viales, se analizan mediante unos moldes para comprimidos y se disponen de tal manera que la ranura de cada comprimido quede en la parte de arriba (para que la incidencia de la radiación sea unidireccional ya que la parte inferior es más plana). Estos moldes también se sitúan en el carrusel rotatorio de muestras.

Los granulados se registran mediante un módulo rotatorio de medida con la cubeta y los registros se deben adquirir uno por muestra (a diferencia de los comprimidos) pero esto permite obtener un resultado más representativo al analizar un área más grande de la muestra.



**Imagen 2: Configuración del equipo NIR. Espectrómetro y automuestreador con moldes para los comprimidos de producción.**



**Imagen 3: Configuración del equipo NIR. Espectrómetro y accesorio para granulados.**

En total se constituyen n clases, cada clase es un grupo de espectros concreto. La razón de registrar no sólo los componentes del producto sino también diferentes fases del proceso de fabricación está relacionada con el hecho de que los espectros de NIR difieren los unos con los otros en relación a diferencias químicas y también físicas (así incluimos toda la variabilidad).

### 5.3. Preparación de las muestras de laboratorio.

El calibrado del modelo de cuantificación debe recoger un rango de concentraciones del 70-130%. Para ello, en el laboratorio se preparan muestras por pesada con diferente contenido de principio activo (que posteriormente deberán ser prensadas). Se prepara una mezcla placebo que contiene los excipientes en las proporciones establecidas por el producto farmacéutico. A esta mezcla placebo se le añade el API también por pesada en una balanza analítica.

Patrones	Cantidad real API (g)	Cantidad real placebo (g)	Total real (g) muestra	% API real (respecto la concentración nominal)
m01	5.3867	4.6280	10.01	69.92
m02	5.3817	4.6275	10.01	69.90
m03	5.7517	4.2923	10.04	74.44
m04	5.7375	4.2912	10.03	74.37
m05	6.0486	3.9629	10.01	78.54
m06	6.0551	3.9547	10.01	78.64
m07	6.3973	3.6207	10.02	83.02
m08	6.3878	3.6223	10.01	82.96
m09	6.7330	3.2921	10.03	87.31
m10	6.7101	3.2971	10.01	87.17
m11	6.9329	3.0831	10.02	89.98
m12	7.0355	2.9704	10.01	91.41
m13	7.2983	2.6907	9.99	94.98
m14	7.3648	2.6347	10.00	95.75
m15	11.5410	3.4629	15.00	100.00
m16	11.5460	3.4615	15.01	100.02
m17	8.0770	1.9261	10.00	104.97
m18	8.0420	1.9718	10.01	104.40
m19	8.3770	1.6415	10.02	108.70
m20	8.4640	1.5483	10.01	109.90
m21	8.6925	1.3154	10.01	112.91
m22	8.6880	1.3137	10.00	112.92
m23	9.0171	0.9850	10.00	117.20
m24	9.0168	0.9822	10.00	117.23
m25	9.3494	0.6507	10.00	121.54
m26	9.3546	0.6536	10.01	121.51
m27	9.6757	0.3262	10.00	125.76
m28	9.6785	0.3262	10.00	125.76
m29	9.9982	0.0000	10.00	130.00
m30	9.9948	0.0000	9.99	130.00

Tabla 2: Preparación de las muestras patrón en un rango de concentración del 70-130%.

Excipientes del núcleo	Cantidad en 1 comprimido (650 mg)	Concentración (%)
Ligante	31	23.8
Desintegrante	27	20.8
Diluyente	62	47.7
Deslizante y lubricante	6	4.6
Lubricante	4	3.1

Tabla 3: Concentración de los excipientes del núcleo.

Con tal de recoger la variabilidad física del proceso, se evalúa la presión de compactación de los comprimidos farmacéuticos. Pequeñas cantidades de las muestras patrón m15 y m16 (que contienen la concentración nominal del producto farmacéutico) se someten a diferentes fuerzas de compactación (medidas en toneladas): 0.25; 0.5; 1.0; 1.5; 2.0 y 2.5 (Esta operación de compactación se lleva a cabo con una prensa Specac de pastillas para el IR). Se obtienen 12 comprimidos y mediante el equipo NIR se registran por las dos caras. El objetivo de esta evaluación es determinar un rango de presiones (mediante métodos quimiométricos de reconocimiento de pautas) asimilable con los comprimidos de producción para prensar muestras patrón de todo el rango de calibrado y así introducirlas en el modelo de cuantificación.

#### 5.4. Método de referencia: HPLC

##### 1. Especialidad farmacéutica en comprimidos recubiertos.

Determinación por un método HPLC de fase reversa, con detección UV (de longitud de onda variable) y cuantificación por el método del estándar externo.

**Columna:** C18 de 5µm, de longitud 125mm y diámetro interno de 4.6mm.

**Fase móvil:** (mezcla de A y B en gradiente)

- A. Disolución A: Metanol (66:34).
- B. Metanol.

Tiempo (min)	Fase móvil A (%)	Fase móvil B (%)
0	100	0
8	100	0
14	70	30
20	70	30

NOTA: La disolución A es Agua purificada: Ácido acético glacial (67:5), siendo el disolvente el sulfóxido de dimetilo

**Post run:** 4 min

**Tiempo del cromatograma:** 20 min.

**Temperatura de la columna:** 40°C

**Flujo:** 1.5ml/min.

**Longitud de onda del detector:** 275nm

**Volumen de inyección:** 10µl

**Preparación de las muestras:** molturar y pesar cada comprimido de manera exacta en un matraz aforado de 100 ml. Añadir 50 ml de dimetil sulfóxido y ultrasonicar 30 min, atemperar y enrasar con el disolvente. Tomar 5 ml y diluir hasta 25 ml con el disolvente. Filtrar a través de una membrana de 0.45 µm.

**Preparación del patrón:** (preparar por duplicado). Pesar aproximadamente 25mg de API de referencia, exactamente pesados, en un matraz de 25 ml. Añadir 10 ml de disolvente. Ultrasonicar 5 min, atemperar y enrasar con disolvente.



## 2. Especialidad farmacéutica en granulado efervescente en sobres.

Determinación por un método HPLC de fase reversa, con detección UV (de longitud de onda variable) y cuantificación por el método del estándar externo.

**Columna:** C18 de 5µm, de longitud 150mm y diámetro interno de 4.6mm.

**Fase móvil:** (mezcla de A y B en proporciones volumétricas 80:20)

- A. 0.5ml de ácido orto-fosfórico al 85% más 340 ml de acetonitrilo y enrasar a 1000 ml con agua. Filtrar a través de un filtro de membrana de 0.45 µm.
- B. Acetonitrilo.

**Tiempo de retención esperado del API:** 8.7 min

**Temperatura de la columna:** temperatura ambiente.

**Flujo:** 2.0ml/min.

**Longitud de onda del detector:** 264nm

**Volumen de inyección:** 20µl

**Preparación de las muestras:** transferir el contenido de un sobre a un matraz de 200ml. Lavar el sobre con 10 ml de metanol, incorporando el metanol en el matraz. Añadir al matraz 20 ml de agua y agitar hasta que termine la efervescencia. Ultrasonicar 20 min, atemperar y enrasar con metanol. Filtrar a través de membrana de nylon de 0.45 µm. (El peso de la muestra se calcula por diferencia de los sobres llenos y vacíos)

**Preparación del patrón:** (preparar por duplicado). Pesar 60mg de API de referencia (±10%), exactamente pesados, en un matraz de 20 ml. Añadir 5 ml de agua y unos 10 ml de metanol. Ultrasonicar 5 min, atemperar y enrasar con metanol.

La fórmula matemática que debe seguirse a fin de obtener, a partir del cromatograma, las concentraciones de las muestras, pudiéndolas, así, introducir las en el modelo de cuantificación NIR como concentraciones reales, es:

$$C_p = \frac{C_{st} \cdot R}{A_{st}} \cdot A_p \cdot \text{Dilución} \cdot \frac{1}{\text{Peso real comprimido o sobre}} \cdot \frac{\text{Peso teórico comprimido o sobre}}{\text{Peso teórico de API}} \cdot 100$$

$C_p$  = concentración de la muestra problema.

$C_{st}$  = concentración del estándar.

$A_p$  = área de la muestra problema.

$A_{st}$  = área del estándar.

R = riqueza del patrón en tanto por uno

## 6. RESULTADOS Y DISCUSIÓN

### 6.1. Modelo de identificación

#### 6.1.1. Especialidad farmacéutica en comprimidos recubiertos

##### Construcción de la biblioteca

Las diferentes clases de la biblioteca espectral comprenden espectros de muestras que pertenecen a un periodo de manufacturación suficientemente amplio (18 meses) como para ser representativos de las fuentes usuales de variación (como suministrador de materias primas, parámetros físicos, proceso de producción...)

La biblioteca espectral consta de varias subbibliotecas que recogen clases que contienen espectros semejantes y que por ello necesitan un paso extra de tratamiento quimiométrico, dado que en el paso de identificación no logran ser discriminados. Los esquemas siguientes muestran el árbol jerárquico de identificación.

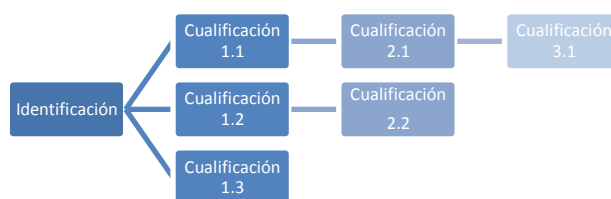


Diagrama 1: Conjunto de biblioteca y subbibliotecas.

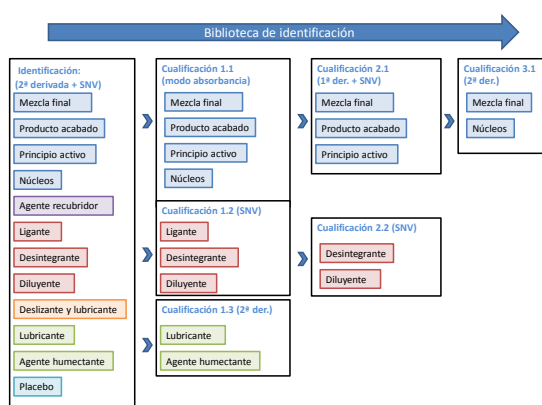


Diagrama 2: Construcción de la biblioteca de identificación.

En cada paso, los espectros que definen una clase generalmente primero se someten a un procesamiento matemático con el objetivo de aumentar la relación señal/ruido y aumentar las diferencias entre clases. Luego a dichos espectros tratados se les aplica un algoritmo de descomposición en componentes principales y se definen las clases mediante cálculos de distancias euclídeas.

	Identification	Qualification 1.1	Qualification 2.1	Qualification 3.1
Spectral range (cm <sup>-1</sup> )	7995.86-3776.15	4451.15-4235.15 4590-4466.58 4898.58-4825.29 5959.29-5646.86	4451.15-4235.15 4894.72-4478.15 5955.43-5473.29	8979.44- 4061.58
Distance	Euclidean	Euclidean	Euclidean	Euclidean
Algorithm	Factorization	Factorization	Factorization	Factorization
Mathematical pre-treatment	2nd derivate (9 smoothing points)+ SNV	Without processing	1st derivative (17 smoothing points) + SNV	2ª Derivate (9 smoothing points)

	Identification	Qualification 1.2	Qualification 2.2
Spectral range (cm <sup>-1</sup> )	7995.86-3776.15	9102.87 – 4003.72	4096.29- 9102.87
Distance	Euclidean	Euclidean	Euclidean
Algorithm	Factorization	Factorization	Factorization
Mathematical pre-treatment	2nd derivative (9 smoothing points)+ SNV	SNV	SNV

	Identification	Qualification 1.3
Spectral range (cm <sup>-1</sup> )	7995.86-3776.15	9018.01-3911.15
Distance	Euclidean	Euclidean
Algorithm	Factorization	Factorization
Mathematical pre-treatment	2nd derivative (9 smoothing points)+ SNV	2nd derivative (9 smoothing points)

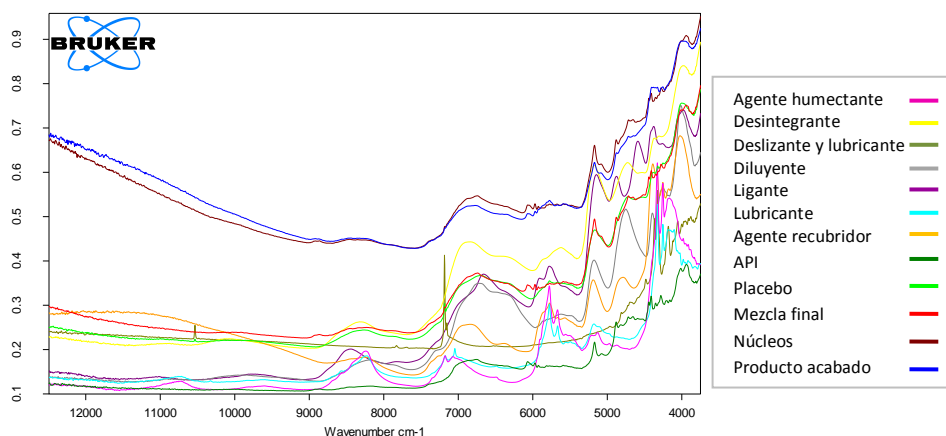
Tabla 4: Parámetros espectrales de la biblioteca de identificación.

Justificación pre-tratamientos y rangos espectrales

Los únicos parámetros a variar son el rango espectral y el pre-tratamiento matemático con el fin de diferenciar las clases (dado que siempre se ha utilizado el mismo algoritmo basado en el cálculo de la distancia euclídeas).

Aunque en la introducción se han dado unas pinceladas acerca de lo que supone la aplicación de cada procesamiento matemático, a priori no puede saberse cuál será el que mejor diferencie las clases. Así que por ensayo y error se prueban metódicamente uno detrás de otro. De igual modo, el uso de un determinado rangos amplios o un conjunto de rangos espectrales es consecuencia de un largo ensayo experimental.

Cabe destacar que pese a que el equipo NIR registra los espectros desde 12500 hasta 3750  $\text{cm}^{-1}$ , no todo ese rango tiene sentido químico: las muestras absorben desde 9200-3700  $\text{cm}^{-1}$  aproximadamente, y el resto suele ser ruido espectral. A priori, siempre se utiliza todo este último



rango, con el objetivo de no perder información. Si esta estrategia no funciona, selecciono las ventanas en las que absorbe el analito que interesa y no lo hace la matriz, o simplemente regiones en las que hay absorción.

Ilustración 10: Espectros de todas las clases de la biblioteca en rango.

El software quimiométrico realiza los cálculos y muestra en el informe de validación interna los valores de los umbrales y desviaciones estándar.

CLASE	Nº espectros	Identificación		Cualificación 1.1		Cualificación 2.1		Cualificación 3.1	
		Nivel de confianza (%)	Umbral	Nivel de confianza (%)	Umbral	Nivel de confianza (%)	Umbral	Nivel de confianza (%)	Umbral
Núcleos	48	99.8	0.017892	100.0	0.074181	100.0	0.13951	99.6	0.53882
Mezcla final	6	97.1	0.020186	99.1	0.059014	99.6	0.064979	99.8	0.14825
Producto acabado	51	100.0	0.086448	99.1	0.046877	99.9	0.38487		
API	3	92.9	0.003419	91.1	0.0053157				
				Cualificación 1.2		Cualificación 2.2			
Diluyente	17	99.5	0.045792	99.1	0.16711	98.9	0.21734		
Desintegrante	17	98.6	0.034525	99.8	0.13554	99.7	0.39017		
Ligante	3	99.2	0.0023395	94.0	0.0094822				
				Cualificación 1.3					
Agente humectante	25	98.9	0.025901	99.0	0.22569				
Lubricante	7	97.4	0.0010979	97.6	0.046016				
Deslizante y lubricante	16	99.3	0.0088875						
Placebo	3	94.2	0.0088486						
Agente recubridor	3	95.1	0.0027322						

Tabla 5: Umbrales y niveles de confianza característicos de cada clase para la identificación y las subbibliotecas.

### Validación interna

La identificación y las cualificaciones son positivas si el valor del hit (que es la distancia entre el centro de la distribución de los espectros de una clase y el espectro a ser identificado) es menor que el umbral establecido para la clase en cuestión.

<i>Fecha de producción</i>	<i>Lote</i>	<i>Identificación</i>	<i>Valor de Hit</i>
Enero 2012	F001	Producto acabado	0,22150
Octubre 2012	F002	Producto acabado	0,10659
Octubre 2012	F003	Producto acabado	0,06939
Octubre 2012	F004	Producto acabado	0,09776
Junio 2013	G701	Producto acabado	0,10018
Junio 2013	G702	Producto acabado	0,17318
Julio 2013	G711	Producto acabado	0,03387

Tabla 6: Resultados de la validación interna. Se muestran los resultados promedio de las distancias para cada lote. Todos los valores de Hit son inferiores al umbral de cualificación 2.1 para la clase de Producto acabado (0,38487).

### Validación externa

Una vez que ya se ha validado internamente la biblioteca espectral (y todas las subbibliotecas) y las clases se diferencian sin ambigüedades, se debe validar la biblioteca con muestras de la misma especialidad farmacéutica que no se han utilizado en la construcción de la misma (que en nuestro caso son comprimidos distintos que pertenecen a los mismos lotes).

<i>Fecha de producción</i>	<i>Lote</i>	<i>Identificación</i>	<i>Valor de Hit</i>
Enero 2012	F001	Producto acabado	0,19627
Octubre 2012	F002	Producto acabado	0,09792
Octubre 2012	F003	Producto acabado	0,06016
Octubre 2012	F004	Producto acabado	0,10525
Junio 2013	G701	Producto acabado	0,12753
Junio 2013	G702	Producto acabado	0,17655
Julio 2013	G711	Producto acabado	0,05216

Tabla 7: Resultados de la validación externa. Especificidad y robustez. Se muestran los resultados promedio de las distancias para cada lote. Todos los valores de Hit son inferiores al umbral de cualificación 2.1 para la clase de Producto acabado (0,38487).

### Validación con comprimidos varios

Se realiza una validación con muestras de productos, que por su semejanza física o química puedan confundirse con el producto del cual se intenta crear un modelo de identificación.

<i>Producto</i>	<i>Fecha de Producción</i>	<i>Hit value</i>	<i>Umbral límite</i>	<i>Etapas de discriminación</i>
Azitromicina 500 mg	Nov-09	0.35125	0.046877	Cualificación 1.1
Terbinafina 250 mg	Dic-09	0.27372	0.046877	Cualificación 1.1
Luminal 0.1 mg	Oct-09	0.31801	0.046877	Cualificación 1.1
Paracetamol 650 mg	Abr-10	0.36245	0.046877	Cualificación 1.1
Diazepan 5mg	Feb-10	0.40242	0.38487	Cualificación 2.1
Digoxina 0.25mg	May-11	0.48152	0.38487	Cualificación 2.1
Furosemida 40mg	Dic-10	0.37846	0.046877	Cualificación 1.1
Gabapentina 600 mg	Ene-10	0.31713	0.046877	Cualificación 1.1

Tabla 8: Resultados de la validación externa con comprimidos varios. La identificación como producto acabado es negativa. Algunos productos se discriminan en el umbral de cualificación 1.1 (0,046877) y otros en la cualificación 2.1 (0,38487).

### 6.1.2. Especialidad farmacéutica en forma de granulado efervescente

El modelo de identificación para este producto ya estaba construido para el equipo NIR del laboratorio. Se debe determinar si es necesaria una modificación de la biblioteca espectral para poder utilizarla en el equipo NIR de la planta de producción.

El archivo del método de identidad se copia al ordenador del equipo planta de producción. Una serie de sobres se analizan por ambos equipos y se comparan los resultados de identidad.

#### Resultados del ensayo de identidad.

El umbral establecido para pertenecer a la clase de producto acabado en el modelo ya construido es de 0.04387. De los sobres analizados por ambos equipos, todos dieron una identificación positiva en el equipo de laboratorio y uno de ellos dio negativa en el de fábrica por lo que se procedió a la revalidación del modelo.

#### Construcción del modelo.

Se añaden los nuevos lotes a la calibración de la biblioteca espectral para que esta recoja toda la variabilidad habitual de las muestras y la variabilidad entre los equipos. El nuevo umbral que define a la clase de producto acabado es ligeramente diferente del antiguo.

<b>Parámetros</b>	<b>Identificación</b>
Rango espectral	3749.1-9002.6 cm <sup>-1</sup>
Distancia	Euclídea
Algoritmo	Estándar (Distancia Euclídea)
Pretratamiento matemático	Normalización vectorial
Nivel de confianza	99.79%
Umbral	0.04453

**Tabla 9: Biblioteca espectral ampliada. Parámetros de la biblioteca.**

#### Validación del modelo de identificación

El modelo de identificación modificado debe validarse internamente para asegurar su consistencia y posteriormente, se ha validado externamente para evitar ambigüedades e identificaciones no correctas (con muestras no utilizadas para la construcción del mismo que abarcan un periodo de fabricación de 1 mes).

También se lleva a cabo una validación externa de granulados varios (para comprobar que la identificación como producto acabado de especialidades farmacéuticas similares es negativa). El procedimiento es análogo al de la validación del producto farmacéutico en comprimidos.

Todas las muestras se identifican correctamente, con todos los valores de hit inferiores al umbral establecido.

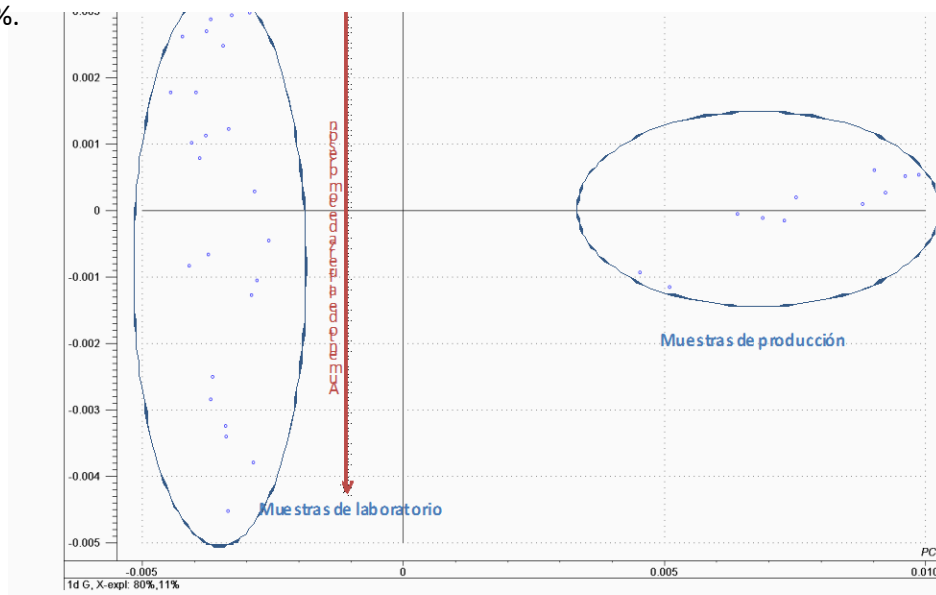
## 6.2. Modelo de cuantificación

### 6.2.1. Especialidad farmacéutica en comprimidos recubiertos.

#### Evaluación de la fuerza de compresión<sup>1</sup>

Se comprimen muestras de concentración nominal de principio activo en un rango amplio de distintas fuerzas. A continuación se registran los espectros de los comprimidos fabricados en el laboratorio.

El gráfico siguiente de scores representa espectros de muestras de producción y las muestras de laboratorio (después de ser sometidos a una primera derivada). El primer componente principal (PC1) corresponde a un 80% de la variabilidad total y el segundo componente (PC2) un 11%.



**Gráfico 1: Gráfico de scores en primera derivada de muestras de laboratorio y muestras de producción. Los gráficos de scores en normalización vectorial o en segunda derivada dan información análoga.**

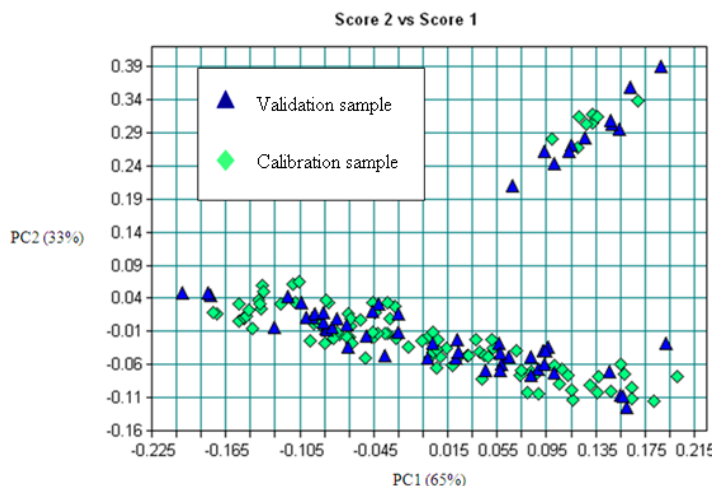
Se puede deducir del gráfico que la mayor fuente de variabilidad, que está asociada al eje de abscisas, es el tipo de muestra (de laboratorio –comprimido fabricado manualmente- o de producción -producto acabado-). La segunda fuente de variabilidad podemos relacionarla con la fuerza de compactación. La flecha roja indica el sentido ascendente de la fuerza de compactación de los comprimidos de laboratorio. La ubicación en el eje de ordenadas de las muestras de producción nos da una idea orientativa de un rango más estrecho de fuerzas de compactación a las cuales someteremos todas las muestras patrón (m01, m02...m30) con el objetivo de crear un conjunto de calibración que contemple toda la variabilidad del proceso de fabricación en un abanico suficientemente amplio de concentraciones nominales de API. El nuevo rango de fuerzas de compactación (medidas en toneladas) es: 1.5, 2.0 y 2.5.

<sup>1</sup>En el apartado de “4.2. Descripción de la creación del modelo de cuantificación” y en “5.3. Preparación de las muestras de laboratorio” de la sección del Procedimiento Experimental se detalla la justificación de la evaluación de la fuerza de compactación, así como la tabla de muestras patrón y el rango de presiones utilizado.

### Creación del conjunto de calibración

El conjunto de todas las muestras es dividido en dos subgrupos, el conjunto de calibración y el de validación. El primero es usado para construir la ecuación del modelo PLS y el segundo para evaluar la capacidad predictiva de dicho modelo.

Los dos subconjuntos deberán contener toda la variabilidad espectral y de concentraciones, esto se consigue mediante un reparto adecuado de muestras de laboratorio preparadas por pesada y muestras de producción analizadas por el método de referencia (HPLC) entre los dos grupos. La distribución de todo el conjunto de muestras en estos dos subconjuntos se realiza mediante un análisis en componentes principales.

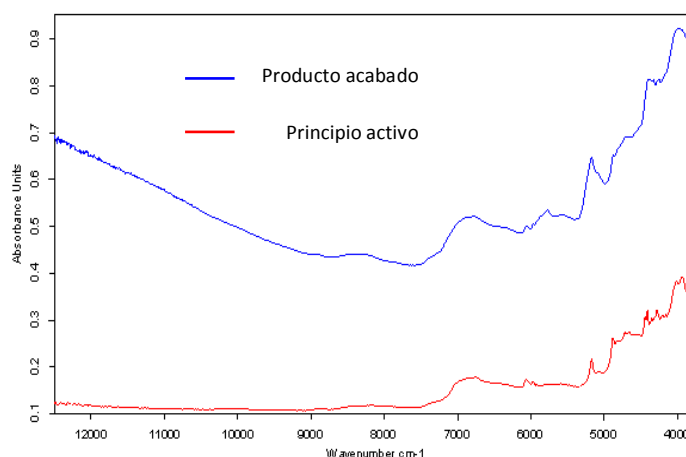


**Gráfico 2:** Selección de los conjuntos de validación y calibración. Gráfico scores de los dos subconjuntos. Los espectros fueron sometido a una segunda derivada (9 puntos de suavizado) en un rango de 9200 a 4000  $\text{cm}^{-1}$ .

### Aplicación del modelo quimiométrico

La obtención de un buen modelo es fruto de un trabajo minucioso, a base de ensayo y error, de ir cambiando las características del mismo (pre-tratamiento matemático, rango o conjunto de rangos, exclusión de determinadas muestras out-liers...).

El modelo escogido posee una amplia región de calibración, es decir, un rango amplio de números de onda (de  $9200 \text{ cm}^{-1}$  a  $4000 \text{ cm}^{-1}$ ), esta estrategia suele proporcionar mayor robustez del modelo frente a cambios inesperados. En este rango aparecen las bandas de absorción del API y se excluyen aquellas que son demasiado débiles o que sólo proporcionan ruido espectral. Así que el modelo de regresión ha sido construido de manera que se asocien cambios de



**Ilustración 11:** Espectros de una muestra de producción de producto acabado y de principio activo en un rango de  $12500\text{-}4000 \text{ cm}^{-1}$ .

concentración a variaciones en la intensidad de las bandas de absorción del ingrediente activo.

Es necesario aplicar un tratamiento espectral que permita incrementar la relación señal/ruido con el fin de mejorarla señal que pertenece al analito.

El pre-tratamiento escogido ha sido una segunda derivada de 9 puntos de suavizado, que permite aumentar las diferencias entre las bandas anchas y solapadas de los espectros, concretamente, minimiza las diferencias de pendiente.

Las siguientes imágenes muestran el rango de números de onda usados para construir el modelo a diferentes concentraciones de principio activo.

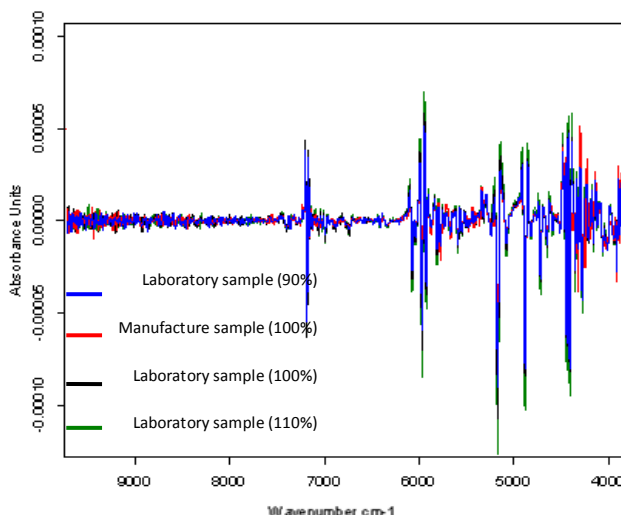
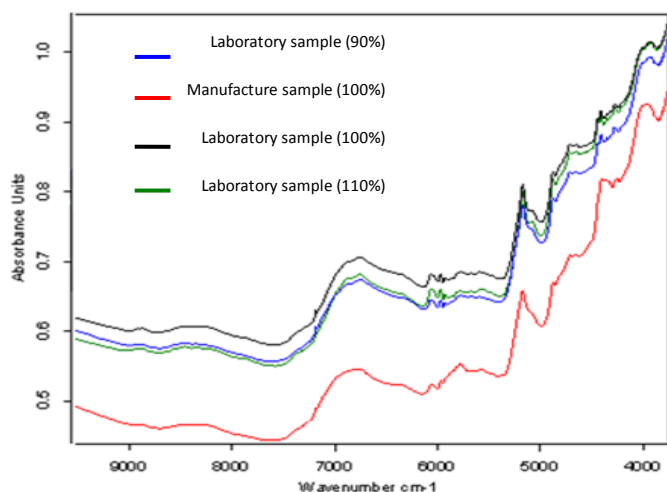


Ilustración 12: Espectros del conjunto de calibración (en unidades de absorbancia)

Ilustración 13: Espectros del conjunto de calibración (en unidades de absorbancia)

### Validación del modelo

La validación externa (mediante el conjunto de validación definido previamente) proporciona los parámetros del modelo de cuantificación, tales como  $R^2$ , nº de factores, bias, RPD, RMSEP... Estos parámetros definen la calidad del modelo y la capacidad que tiene para predecir las muestras.

El número de componentes PLS necesarios para asegurar una buena capacidad predictiva depende de varios factores, incluyendo la concentración nominal de principio activo en el producto, las propiedades químicas o la complejidad del sistema estudiado. La elección de un número correcto de factores se basa en el cálculo de RMSEP (Root Mean Square of Prediction).

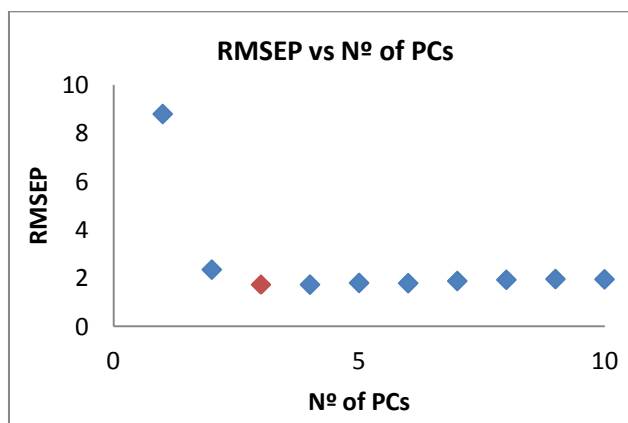


Gráfico 3: Gráfico RMSEP según el número de componentes PLS.

El menor valor de la curva de RMSEP indica el número recomendado de factores. El siguiente gráfico muestra los valores RMSEP en función del número de componentes PLS para el modelo. Para este modelo se han usado tres componentes PLS; con estos factores el error es 1.74. La adición de más componentes PLS no mejora la capacidad predictiva del modelo.



El siguiente gráfico proporciona información adicional, concretamente, se muestran los valores estadísticos de  $R^2$  según el número de componentes PLS. El mayor valor de  $R^2$  (98,63) corresponde al número apropiado de componentes PLS del modelo de cuantificación.

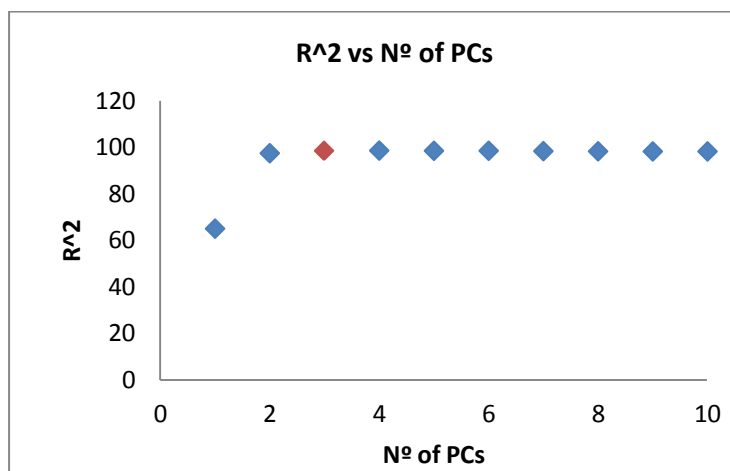


Gráfico 4: Gráfico de  $R^2$  según el número de componentes PLS.

Nº of PCs	RMSEP	R <sup>2</sup>
1	8.78	65.03
2	2.34	97.52
3	1.74	98.63
4	1.72	98.65
5	1.79	98.55
6	1.78	98.56
7	1.87	98.41
8	1.92	98.33
9	1.95	98.28
10	1.94	98.29

Tabla 10: Valores de RMSEP y  $R^2$  en función del número de componentes PLS.

#### Parámetros estadísticos del modelo desarrollado

Se debe documentar la validez del modelo PLS desarrollado mediante el cálculo de parámetros estadísticos (tales como exactitud, precisión, linealidad...) con tal de poder poner en uso el método de cuantificación y hacerlo válido respecto al método de referencia establecido.

**Selectividad.** Describe la habilidad de un procedimiento analítico para diferencias entre varias sustancias en la muestra. En el caso que nos ocupa, la selectividad se demuestra mediante el uso de la biblioteca espectral, obteniendo un umbral inferior al de la clase determinada por el producto acabado.

El modelo de identificación validado permite que se aplique el modelo de calibración una vez que la muestra se identifica correctamente.

**Linealidad.** Es la capacidad, dado un rango determinado, de obtener resultados que sean directamente proporcionales a la concentración de analito en la muestra.

Se representa los resultados de ambos métodos en los dos ejes de una línea de regresión, siendo cada punto una medida realizada por cada uno de los dos métodos. Puesto que existen errores aleatorios, habrá una cierta desviación del resultado ideal, pero para aceptar el nuevo método se espera que el intervalo de confianza de la ordenada en el origen incluya el cero y que el intervalo de confianza de la pendiente incluya el 1.

En base a los resultados obtenidos en la regresión de comparación del método NIR propuesto y el método de referencia HPLC, los valores de los coeficientes estimados y sus correspondientes intervalos de confianza para un 95% de significación, ambos métodos proporcionan resultados aceptables y consistentes a lo largo de todo el rango de concentraciones estudiado, sin detectarse ninguna muestra anómala. Ver tabla 16

Según las guías de la EMA (Agencia Europea del Medicamento), para que la linealidad del método NIR sea válida respecto al método de referencia, el SEP (Standard Error of Prediction) debe ser menor que 1.4 veces el error estándar de laboratorio del método de referencia ( $SEL_{HPLC}$ ). [17] El error estándar de laboratorio (SEL) del método de referencia (HPLC) se ha calculado a partir de la comparación de los resultados seis análisis llevados a cabo por dos analistas en diferentes días.

$$SEL_{HPLC} = \sqrt{\sum \text{Residuales}^2 / n^{\circ} \text{ de muestras}}$$

% API por HPLC		
Muestra	Analista 1	Analista 2
PR1	101.8	98.4
PR2	100.3	99.1
PR3	100.0	98.5
PR4	100.2	97.3
PR5	100.3	97.4
PR6	100.4	98.6
$SEL_{HPLC}$	2.428	

Tabla 11: Evaluación del error estándar de laboratorio para el método de referencia.

Tal y como recomienda la guía EMA el  $SEP_{NIR}$  1.74 es menor que  $1.4 \times SEL_{HPLC}$  (2.43%). El modelo PLS satisface los criterios de aceptación para la linealidad.

**Exactitud:** expresa el grado de concordancia entre el valor predicho y el valor aceptado o un valor verdadero establecido por convenio o el valor de un método de referencia. Para evaluar la exactitud se usan materiales de referencia certificados y se compara el método analítico desarrollado con un método de referencia.

Tipo de muestra	Lote de la muestra	Valor de referencia (% del valor nominal)	Valor NIR (% del valor nominal)	Residual
Producción	F002	104.0	101.6	2.4
Producción	F002	103.3	101.8	1.5
Producción	F003	101.5	98.8	2.7
Producción	F004	100.5	102.9	-2.5
Producción	F004	102.3	99.0	3.4
Producción	F004	101.3	98.8	2.5

Tabla 12: Evaluación de la exactitud. Resultados de cuantificación de API por ambos métodos.

La exactitud se documenta mediante la aplicación de un test t de datos aparejados. Ver tabla 16. Los resultados demuestran que no se observan diferencias significativas entre ambos métodos para un nivel de significación del 95%.

**Precisión:** expresa el grado de dispersión entre series de múltiples medidas de la misma muestra homogénea bajo unas determinadas condiciones. La precisión de un método analítico se expresa normalmente como la varianza.

- **Repetibilidad:** es la precisión bajo las mismas condiciones operativas en un corto intervalo de tiempo.

<b>Determinación</b>	<b>Valor NIR (% del valor nominal)</b>
1	98.9
2	99.9
3	100.1
4	98.7
5	99.9
6	99.4

Tabla 13: Evaluación de la repetibilidad. Resultados del análisis de un sólo comprimido.

El valor obtenido de la desviación estándar relativa (RSD) demuestra la adecuada repetibilidad del método NIR propuesto. Ver tabla 16

- **Precisión intermedia:** expresa las variaciones dentro de un mismo laboratorio. Se evalúa en diferentes días con distintos analistas.

<b>Operador</b>	<b>Día 1</b>	<b>Día 2</b>	<b>Día 3</b>
A	101.2	102.0	103.2
B	99.1	99.8	103.9

Tabla 15: Resultados en % de los análisis de una misma muestra en días diferentes por dos analistas distintos.

<b>Día</b>	$F_{exp}$	5.0
	$F_{tab}$	19.0
<b>Operador</b>	$F_{exp}$	1.8
	$F_{tab}$	18.5

Tabla 14: Valores de  $F_{exp}$  y  $F_{tab}$  para los factores Día y Operador.

El análisis de la varianza de dos vías para un 95% de nivel de significación revela que no se aprecian diferencias significativas entre días y operadores. Ni el factor día ni el factor operador introducen diferencias. Las conclusiones de los resultados del análisis de varianza permiten afirmar que la precisión intermedia del método NIR propuesto es adecuada. Ver tabla 16.

**Robustez:** es la capacidad de un método analítico de producir resultados válidos en presencia de cambios inesperados de las condiciones externas (diferentes lotes, cambios de suministrador, estabilidad a lo largo del tiempo...).

Para evaluar la robustez se han analizado 6 muestras (que comprenden un periodo de fabricación de 1 mes) por el método NIR creado y por el de referencia y se ha llevado a cabo un test t de datos emparejados. Los resultados son los mismos que los de la evaluación de la exactitud, dado que se carece de más muestras. La puesta en marcha del nuevo método de cuantificación NIR creado exige una revisión de la robustez, analizando más muestras por ambos métodos para garantizar que el método es proporciona resultados válidos a lo largo del tiempo. Ver tabla 16.

**Tabla resumen de los parámetros de calidad del modelo de cuantificación NIR**

Aspecto	Procedimiento	Modelo de laboratorio	
<b>Linealidad</b>	[NIR] = a + b · [REF]	Muestras	58
		Rango de concentraciones	70-130% concentración nominal de API
		b	0.99± 0.03
		a	0±3
		Correlación	0.99
<b>Exactitud</b>	Test t de daos apareados de valores NIR y valores de REF de lotes de producción	Muestras	6
		Valor medio de los residuales	1.6
		SD	2.10
		t <sub>exp</sub>	1.92
		t <sub>crit</sub>	2.57
<b>Repetibilidad</b>	Muestras de producción analizadas seis veces por el mismo operador	Promedio valores NIR	99.5
		SD	0.6
		RSD	0.6
<b>Precision Intermedia</b>	Muestras de producción analizadas en tres días por dos operadores	Promedio valores NIR	101.5
		SD	1.88
		RSD	1.86
<b>Robustez</b>	Test t de datos apareados de valores de NIR y valores de REF de lotes de producción analizadas en un periodo de 1 mes.	ANOVA de dos factores	No hay efecto significativo de día u operador
		Muestras	6
		Valor medio de los residuales	1.6
		SD	2.10
		t <sub>exp</sub>	1.92
t <sub>crit</sub>	2.57		

Tabla 16: Resultados obtenidos en la validación del modelo de calibración construido para cuantificar API en la especialidad farmacéutica de comprimidos recubiertos (expresados en %).

### 6.2.2. Especialidad farmacéutica en forma de granulado efervescente.

El modelo de cuantificación para este producto ya estaba construido para el equipo NIR del laboratorio. Se debe determinar si es necesaria una transferencia de calibración para poder utilizar dicho modelo en el equipo NIR de la planta de producción.

El archivo del método de cuantificación se copia al ordenador del equipo planta de producción. Una serie de sobres se analizan por ambos equipos y se comparan los resultados de contenido en % de principio activo.

#### Resultados del ensayo

Se realiza un test t de datos emparejados (95% de confianza, de dos colas) para evaluar si existen diferencias significativas entre los resultados cuantitativos de los dos equipos para una serie de sobres.

Valor medio residuales	1.6	Los resultados del test de datos emparejados muestran que existen diferencias significativas entre los resultados de cuantificación de los 2 equipos a un nivel de significación del 95%. El valor experimental del parámetro estadístico t es mayor que el valor crítico de dicho parámetro.
Desviación estándar	5.6	
Número de muestras	54	
Valor experimental de t	2.1	
Valor crítico de t (95%, 53 g.d.l.)	2.0	

**Tabla 17: Resultados de la comparativa de datos apareados para los equipos de producción y de laboratorio.**

Los resultados del test indican que es necesario revalidar el modelo de cuantificación.

#### Creación del conjunto de calibración.

De manera análoga a la creación del modelo de cuantificación para la especialidad farmacéutica en comprimidos recubiertos, se divide el conjunto de muestras en dos subgrupos: el conjunto de validación y el de calibración. Para las muestras de producción (que son estos nuevos sobres) se determinó el valor de referencia de la concentración nominal de API por HPLC, y para las muestras de laboratorio (ya estaban creadas para el modelo de cuantificación del laboratorio) el mismo valor se determinó por pesada. Los conjuntos de calibración y validación están constituidos por muestras de laboratorio y muestras de producción.

El conjunto de calibración abarca un rango de concentraciones de 73,7% a 123,1% de API (respecto a la concentración nominal). Las muestras contenidas en cada conjunto fueron seleccionadas mediante un análisis en componentes principales.

### Aplicación del modelo quimiométrico

El rango de trabajo para el método cuantitativo es:  $9002.6-7999.7\text{cm}^{-1}$ ;  $7502.1-6996.9\text{cm}^{-1}$ ;  $6102.0-5596.7\text{cm}^{-1}$  y  $4501.3-4196.6\text{cm}^{-1}$ . Se construye el modelo de regresión con aquellos números de onda que proporcionan información sobre el principio activo a cuantificar.

Los espectros han sido sometidos a una derivada de primer orden con ventana móvil de 13 puntos más una normalización vectorial (SNV). El pre-tratamiento espectral permite aumentar la relación señal/ruido y por consiguiente mejorara la señal correspondiente al analito.

Posteriormente se aplicó un método de regresión por mínimos cuadrados parciales (PLSR).

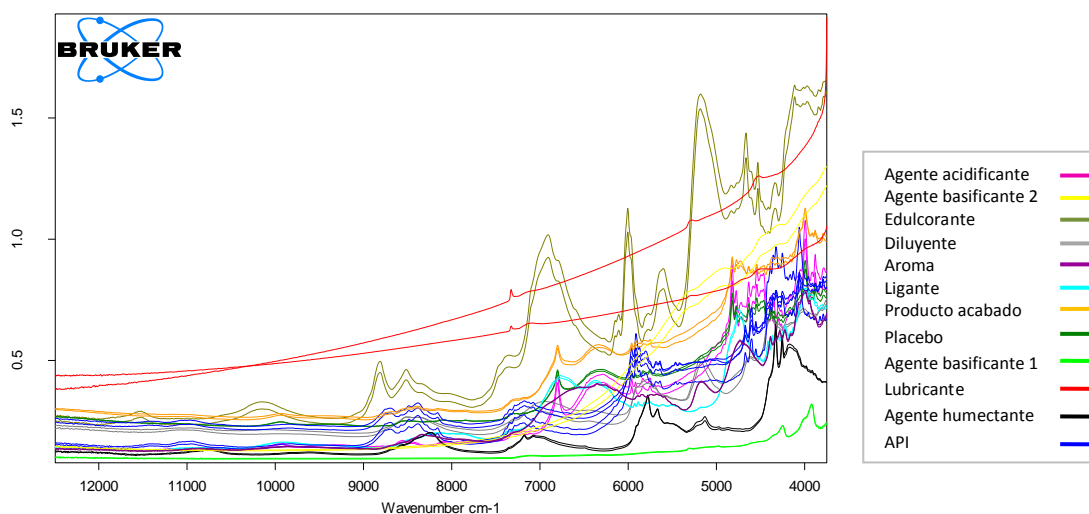


Ilustración 14: Comparación de los espectros de las materias primas usadas en la fabricación de la especialidad farmacéutica en granulado efervescente.

### Validación del modelo

El número de componentes PLS empleados para la construcción del modelo de calibración han sido dos 7 componentes PLS, con estos factores el coeficiente de correlación,  $R^2$ , es de 99.1 y el error en predicción del 1.0%, la adición de más componentes PLS no mejora estadísticamente la capacidad predictiva del modelo PLSR.

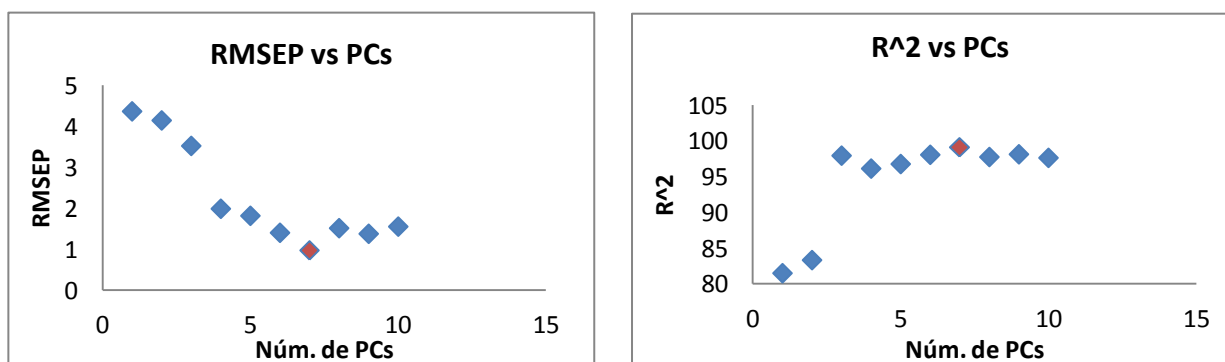


Gráfico 5: Root Mean Standard Error of Prediction vs núm de PCs (Izquierda) y Coeficiente de la regresión al cuadrado vs el núm de PCs.

## Parámetros estadísticos del modelo desarrollado

Tabla resumen de los parámetros de calidad del modelo de cuantificación NIR

Aspecto	Procedimiento	Modelo de laboratorio	
<b>Linealidad</b>	[NIR] = a + b · [REF]	Muestras	19
		Rango de concentraciones	77.6-122.8% conc. nominal de API
		b	0.99± 0.05
		a	1±5
		Correlación	0.99
<b>Exactitud</b>	Test t de daos apareados de valores NIR y valores de REF de lotes de producción	Muestras	13
		Valor medio de los residuales	0.2
		SD	1.1
		t <sub>exp</sub>	0.6
		t <sub>crit</sub>	2.2
<b>Repetibilidad</b>	Muestras de producción analizadas seis veces por el mismo operador	Promedio valores NIR	97.1
		SD	1.8
		RSD	1.8
<b>Precision Intermedia</b>	Muestras de producción analizadas en tres días por dos operadores	Promedio valores NIR	97.4
		SD	2.0
		RSD	2.0
		ANOVA de dos factores	No hay efecto significativo de día u operador
<b>Robustez</b>	Test t de datos aparejados de valores de NIR y valores de REF de lotes de producción analizados en un periodo de 1 mes.	Muestras	13
		Valor medio de los residuales	0.2
		SD	1.1
		t <sub>exp</sub>	0.6
		t <sub>crit</sub>	2.2

Tabla 18: Resultados obtenidos en la validación del modelo de calibración construido para cuantificar API en la especialidad farmacéutica en forma de granulado efervescente (expresados en %).

En esta tabla, equivalente a la mostrada para el modelo de cuantificación del producto en comprimidos recubiertos, se muestran descritos los parámetros estadísticos para el modelo de cuantificación de la especialidad farmacéutica en forma de granulado efervescente. El procedimiento experimental y la estrategia de cálculo se han desarrollado de manera semejante a la descrita en la sección anterior.

Los valores de los parámetros estadísticos son buenos, de modo que el modelo de cuantificación es válido para su aplicación.

## 7. CONCLUSIONES

En base al trabajo realizado se constata que la creación de modelos de identificación y cuantificación mediante espectroscopia NIR puede sustituir a otros métodos de análisis.

Los espectros NIR presentan bandas anchas y solapadas, de modo que una identificación directa por comparación es inviable. Los tratamientos quimiométricos, con previo filtrado de señales, permiten tratar los espectros de modo que puedan discriminarse unos de otros. Concretamente, en este trabajo, se ha aplicado la técnica de reducción de variables PCA basada en distancias euclídeas para la creación de los modelos de identificación o bibliotecas espectrales.

La información química se extrae de los espectros NIR, pero la información física también está incluida en ellos, permitiendo, esto último, la creación de modelos que describan mejor toda la variabilidad del proceso de producción.

PLSR es un método de regresión multivariable (basado en una reducción de variables) que relaciona cambios en el espectro con cambios en la concentración. En este trabajo se ha utilizado PLSR para poder desarrollar modelos de cuantificación a partir de los espectros de muestras de producción y muestras de laboratorio (prensadas en un determinado rango de presiones) y mediante valores verdaderos obtenidos por el método de referencia establecido.

Los dos métodos NIR desarrollados en este trabajo han sido validados con el fin de determinar si permiten una cuantificación suficientemente exacta y precisa de principio activo en las especialidades farmacéuticas evaluadas. Con ese propósito se ha evaluado la linealidad, exactitud, repetibilidad, selectividad, precisión intermedia y robustez del método siguiendo las guías de la Agencia Europea del Medicamento.

En base a los satisfactorios resultados obtenidos en la validación de ambos métodos, éstos, una vez aprobados por la EMA, pueden ser aplicados como métodos absolutos para el análisis de las especialidades farmacéuticas estudiadas.



## 8. BIBLIOGRAFÍA

1. D. A. Skoog, F. J. Holler, T. A. Nieman. *"Principios de análisis instrumental"*. (2001), 5ª ed., McGraw-Hill.
2. J.C.Miller, J. N. Miller. *"Estadística y Quimiometría para Química Analítica"*. Prentice Hall. 4ª Ed. (2002).
3. Ramos Ramis, Guillermo; García Álvarez Coque, Mª Celia. *"Quimiometría"*. Editorial Síntesis (2001).
4. J. Coello, S. Maspocho. *"Calibración Multivariable. En Temas avanzados de Quimiometría"*. M. Blanco, V. Cerdà Eds. (2007). Universitat de les Illes Balears. Palma.
5. *"Multivariate Data Analysis for Dummies"*, .CAMO Software Special Edition. Published by John Wiley & sons , Ltd. (2012)
6. Conzen, Jörg-Peter. *"Multivariate Calibration. A practical guide for developing methods in the quantitative analytical chemistry"*. 2<sup>nd</sup> Edition(2006) Bruker Optic GmbH
7. Peguero Gutiérrez, Anna. Tesis Doctoral. *"La espectroscopia NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabado"*. Programa de doctorado de Química. Departamento de Química. Facultad de Ciencias UAB (2010).
8. Castillo Martínez, Miguel. Tesis doctoral. *"Aplicación de la Espectroscopia NIR al Control Analítico de Procesos de la Industria"* Programa de doctorado de Química. Facultad de Ciencias UAB (2007).
9. Randall D. Tobias, *"An Introduction to Partial Least Squares Regression"*. SAS Institute Inc., Cary, NC.
10. Svante Wold. Michael Sjöström, Lennart Eriksson. *"PLS- regression: a basic tool of chemometrics"*. Chemometrics and Intelligent Laboratory Systems 58 (2001) 109-130
11. Saikat Maitra und Jun Yan. *"Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression"*. Casualty Actuarial Society, 2208 Discussion Paper Program
12. John Shlens. *"A Tutorial on Principal Component Analysis. Derivation. Discussion and Singular Value Decomposition"* (2003)
13. M.Blanco and M.A. Romero. *"Near infrared libraries in the pharmaceutical industry: a solution for identity confirmation."* The Analyst, journal nº126, 2001.

14. M. Blanco; A. Peguero. *“Analysis of pharmaceuticals by NIR spectroscopy without a reference method”*. *Trends in Analytical Chemistry*, Vol. 29, Nº10, 2010.
15. M. Blanco; M.A. Romero; M. Alcalà. *“Strategies for constructing the calibration set for a near infrared spectroscopic quantitation method”*. *Talanta* 64 (2004), 597-602 (Elsevier).
16. Moroto, Alicia; Riu, Jordi; Boqué, Ricard; F. Rius, Xavier; *“Validación de métodos analíticos. Trazabilidad incertidumbre. En Temas avanzados de Quimiometria”* M. Blanco, V. Cerdà Eds. (2007). Universitat de les Illes Balears. Palma.
17. Guidelines for the Development and Validation of Near Infrared (NIR) spectroscopic methods. (2012) Released by EMA (European Medicines Agency).  
[http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2012/02/WC500122769.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/02/WC500122769.pdf)

