



**SEKISUI**

CALIBRACIÓN Y VALIDACIÓN DE UN MÉTODO  
POR ESPECTROSCOPIA NIR PARA LA MEJORA EN  
LA CUANTIFICACIÓN DEL GRADO DE HIDRÓLISIS  
EN MUESTRAS DE ALCOHOL POLIVINÍLICO

**Patrick Iriart Martínez**

Asignatura: Trabajo de fin de grado

Curso: 2020/2021

Tutor académico: Dr. Luca Guerrini

Empresa: Sekisui Specialty Chemicals

Tutor de empresa: Dr. Iñaki Iragorri Sainz



## Agradecimientos

Transmitir mi más sincero agradecimiento a todos aquellos que me han ayudado a lo largo de esta etapa.

En primer lugar, a mi tutor en la empresa, el Doctor Iñaki Iragorri, a la líder de laboratorio, Margarita Negrete y a mi tutor académico, Luca Guerrini, por su ayuda en la planificación, información y organización en este Trabajo de Fin de Grado.

En segundo lugar, a mi familia, mi madre, mi padre y mis hermanos. A mis amigos y a mi pareja Andrea que han estado a lo largo de toda mi carrera apoyándome en todo momento y animándome a seguir adelante.

También, expresar mi más sentido agradecimiento a los trabajadores de la empresa SEKISUI SPECIALTY CHAMICALS, Mari, Sandra, Andrea, Helena, Joan, Elena y Pedro por acogerme dentro de su laboratorio y hacerme sentir como en casa. Después de este período de investigación escribo este apartado de agradecimientos para finalizar mi TFG, TFM o Tesis. Sin duda, ha sido un período de aprendizaje científico y personal.

Desarrollar este estudio ha tenido un gran impacto en mi persona y es por eso por lo que me gustaría agradecerse a todas aquellas personas que me han apoyado durante este proceso.

A todos ellos, mil gracias

# Índice

\_Toc75112075

|  |    |
|--|----|
| 1.- Resumen .....  | 1  |
| 1.- Abstract.....  | 1  |
| 2.- Objetivos.....   | 2  |
| 3.- Introducción .....   | 3  |
| 3.1. Síntesis del alcohol polivinílico .....                                 | 4  |
| 3.1.1. Síntesis de acetato de polivinilo para alcohol polivinílico.....      | 5  |
| 3.1.2. Síntesis de alcohol polivinílico.....                                 | 6  |
| 3.2. Método de referencia.....   | 6  |
| 3.3. Near Infrared (NIR).....  | 7  |
| 3.4. Quimiometría.....   | 8  |
| 3.4.1 Introducción.....  | 8  |
| 3.4.2 Métodos de selección de outliers.....                                  | 9  |
| 3.4.3 Métodos de pretratamiento para la mejora de la señal instrumental..... | 10 |
| 3.4.4 Técnicas de calibrado.....   | 12 |
| 3.4.5 Validación de la recta .....   | 13 |
| 3.5 Ventajas de la utilización del NIR.....                                  | 14 |
| 4. Parte experimental.....   | 15 |
| 4.1 Diagrama de Gantt PVOH escamas y PVOH producto final.....                | 15 |
| 4.2 Adquisición de espectros para el modelo multivariante .....              | 17 |
| 4.3 Selección de muestras.....   | 18 |
| 4.4 Desarrollo de los modelos de calibrado multivariante .....               | 21 |
| 4.5 Validación del modelo de calibrado multivariante .....                   | 25 |
| 5. Resultados y discusión .....  | 28 |
| 6. Conclusión .....  | 33 |
| 6. Conclusion .....  | 34 |
| 7. Bibliografía.....   | 36 |

## 1.- Resumen

En la industria se ha mostrado especial interés en la técnica NIRS (espectroscopia de infrarrojo cercano) debido a su carácter no destructivo, al consumo mucho menor de disolventes y a la reducción de tiempo de análisis.

Usar NIRS para análisis cuantitativos requiere de la aplicación de técnicas de calibración multivariable, en este trabajo utilizaremos la técnica PLS (regresión parcial por mínimos cuadrados).

El objetivo de este trabajo es construir un nuevo modelo de calibración para determinar el grado de hidrólisis en muestras de alcohol polivinílico en distintas fases del proceso, con tal de optimizar el modelo utilizado anteriormente en el laboratorio.

## 1.- Abstract

NIRS (near infrared spectroscopy) has been of particular interest in industry due to its non-destructive nature, much lower solvent consumption and reduced analysis time.

Using NIRS for quantitative analysis requires the application of multivariate calibration techniques, in this work, we will use the PLS (partial least squares regression) technique.

The aim of this work is to build a new calibration model to determine the degree of hydrolysis in polyvinyl alcohol samples at different stages of the process, in order to optimize the model previously used in the laboratory.

## 2.- Objetivos

Objetivo principal:

- Calibración y validación de dos métodos cuantitativos vía espectroscopia de infrarrojo cercano (NIR), uno para el análisis de las escamas (fase de proceso) y otro para el producto final, que permitan la determinación del grado de hidrólisis en muestras de alcohol polivinílico.

Objetivos específicos:

- Mejora del proceso de análisis del grado de hidrólisis de una muestra de alcohol polivinílico. Con esta optimización se reducirá el tiempo y mejorará exactitud de la medida.

### 3.- Introducción

Este proyecto ha sido realizado en Sekisui Specialty Chemicals situado en el polígono sur de Tarragona, el polígono petroquímico más importante del sur de Europa. En esta planta química se fabrica alcohol polivinílico y copolímeros de este. Sekisui Specialty Chemicals lleva produciendo polímeros y copolímeros de alcohol polivinílico de alta calidad desde 1959, aunque se instaló en Tarragona el año 1976.

El alcohol polivinílico Selvol (PVOH) es un polímero soluble en agua no tóxico y respetuoso con el medio ambiente. Este polímero sería el resultado de polimerizar el etenol que no existe (ya que es la forma tautómera inestable del acetaldehído), en su lugar se prepara por alcoholisis (total o parcial) de los grupos acetato en medio básico. Las propiedades de este polímero se pueden modular controlando su grado de hidrólisis. Sus aplicaciones entre otras son:

- Adhesivo en la producción de cartón.
- Agente de hilado de fibras.
- Bolsas hidrosolubles.
- Agente desmoldeante.
- Aditivo para cementos cola.
- Excipiente en farmacia.
- Agente emulsionante para polimerizaciones en emulsión.

A causa de lo comentado anteriormente es muy importante conocer el grado de hidrólisis del polímero. En esta empresa la determinación del grado de hidrólisis se realiza mediante una valoración por retroceso del exceso de ácido sulfúrico con un valorador potenciométrico como método primario, pero durante mi estancia en prácticas en la empresa vamos a tratar de optimizar un método para cuantificar el grado de hidrólisis a partir de NIR para un tipo de producto en concreto. Su uso generalizado se debe principalmente a que permite realizar análisis cualitativos y cuantitativos de multicomponentes en muestras, con un mínimo de preparación. Esta metodología, además, se caracteriza por ser no destructiva, rápida, no emplear reactivos químicos,

disminuir el error del operador y requerir menos mano de obra que los métodos tradicionales empleados en el laboratorio. Otra de las grandes ventajas de la espectrofotometría NIR es su versatilidad ya que presenta varias maneras de analizar la muestra. Sin embargo, se debe tener presente que es un método secundario, lo cual significa que debe ser calibrado en base a valores primarios de referencia, obtenidos mediante otras metodologías. De igual manera sus respuestas no presentarán mayor exactitud que la de los métodos primarios empleados [1].

### 3.1. Síntesis del alcohol polivinílico [2]

El alcohol polivinílico se produce comercialmente mediante la hidrólisis del acetato de polivinilo, ya que el alcohol vinílico monomérico no puede obtenerse en cantidades y pureza que hagan viable la polimerización por radicales libres. Puesto que es la forma tautómera inestable del acetaldehído.

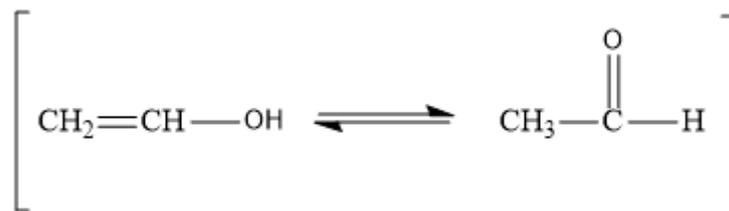


Ilustración 1. Tautomería ceto-enólica del alcohol vinílico monomérico.

En la ilustración número 2 podemos ver el diagrama de bloques del proceso de síntesis del alcohol polivinílico.

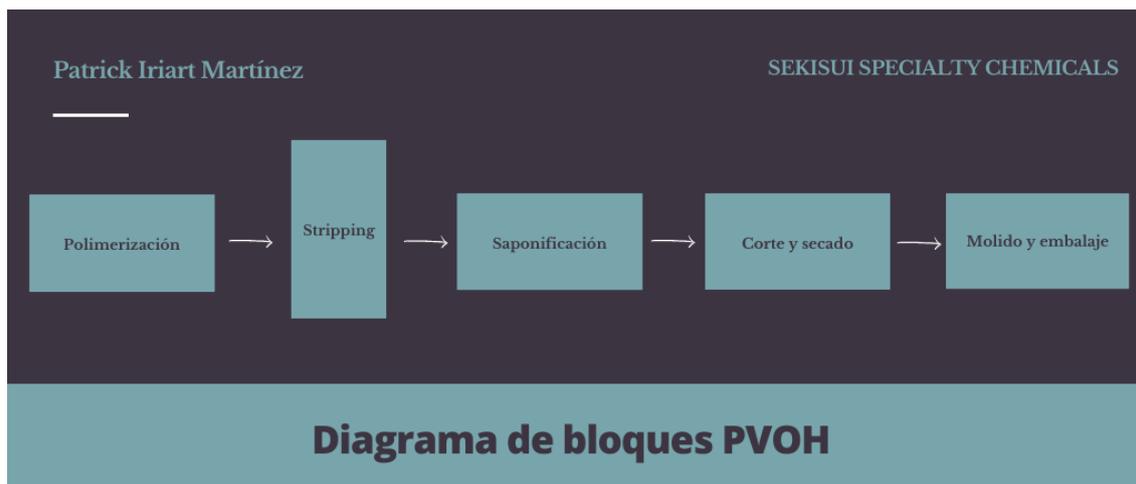


Ilustración 2. Diagrama de bloques del proceso de la síntesis del alcohol polivinílico

### 3.1.1. Síntesis de acetato de polivinilo para alcohol polivinílico

La síntesis del acetato de polivinilo se lleva a cabo a partir de la polimerización radicalaria del monómero acetato de vinilo, también conocido como VAM, con ayuda de un iniciador. Generalmente se utiliza la polimerización en solución con metanol, ya que el polímero resultante se puede hidrolizar directamente. La reacción de polimerización es exotérmica, así que necesitaremos un condensador de reflujo acoplado al reactor.

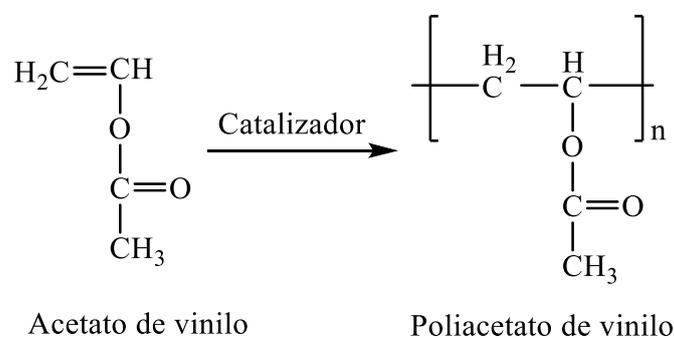


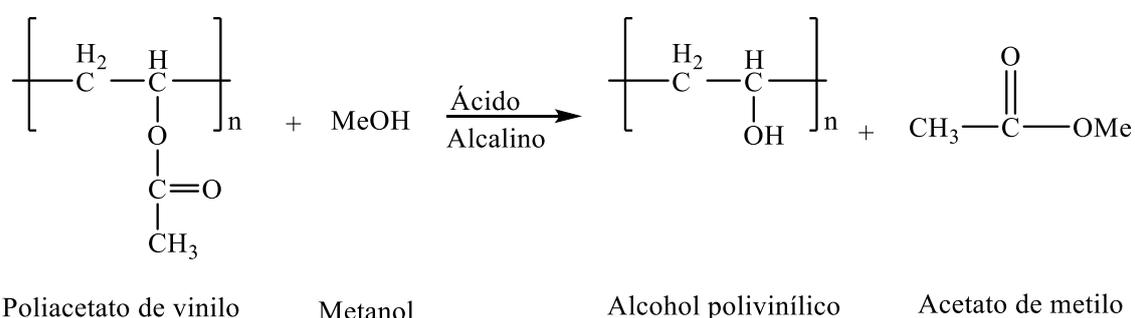
Ilustración 3. Polimerización radicalaria del monómero acetato de vinilo

Dado que la reacción no puede llevarse a cabo por completo, el acetato de vinilo residual se elimina mediante una columna de extracción. Se utilizan vapores de metanol para eliminar el acetato de vinilo residual de la solución de acetato de polivinilo. El acetato de vinilo y el metanol procedentes de la parte superior de la columna de stripping pueden enviarse a una sección de recuperación o reciclarse directamente al reactor.

### 3.1.2. Síntesis de alcohol polivinílico

Después de limpiar el poliacetato de vinilo se introduce dentro del saponificador (reactor).

Industrialmente, la reacción más importante es la de transesterificación, en la que se añade una pequeña cantidad de ácido o base en cantidades catalíticas para primar el cambio de éster. La reacción suele denominarse "saponificación", de ahí el nombre del reactor.



*Ilustración 4. Reacción de transesterificación de la síntesis del alcohol polivinílico*

Una vez realizada la reacción pasamos a la etapa de cortado y secado. Donde el gel resultante de la saponificación es cortado y secado (momento del proceso donde cogemos la muestra a la que llamamos escamas). Finalmente transportamos estas escamas a una moladora y obtenemos el producto final.

### 3.2. Método de referencia

#### **Determinación del grado de hidrólisis del PVOH húmedo**

La determinación del grado de hidrólisis se realiza hidrolizando los grupos acetilo residuales, es decir hidrolizando lo máximo posible el producto, esto se lleva a cabo añadiendo NaOH más calor. Una vez hidrolizada la muestra se añade una cantidad específica de ácido sulfúrico que reacciona con el NaOH sobrante y por último valoramos el ácido sulfúrico no reaccionado con NaOH factorizado.

### 3.3. Near Infrared (NIR)

La región espectral del infrarrojo cercano se extiende desde el extremo superior de longitudes de onda de la región visible, en alrededor de 770 nm a 2500 nm (de 13 000 a 4000  $\text{cm}^{-1}$ ). Las bandas de absorción en esta zona son sobretonos o combinaciones de las bandas vibracionales de tensión fundamentales que se producen en la región de 3000 a 1700  $\text{cm}^{-1}$ . Por lo general, los enlaces afectados son C-H, N-H, y O-H. Puesto que se trata de sobretonos o bandas de combinación, sus absorptividades molares son pequeñas y los límites de detección son del orden de 0.1%.

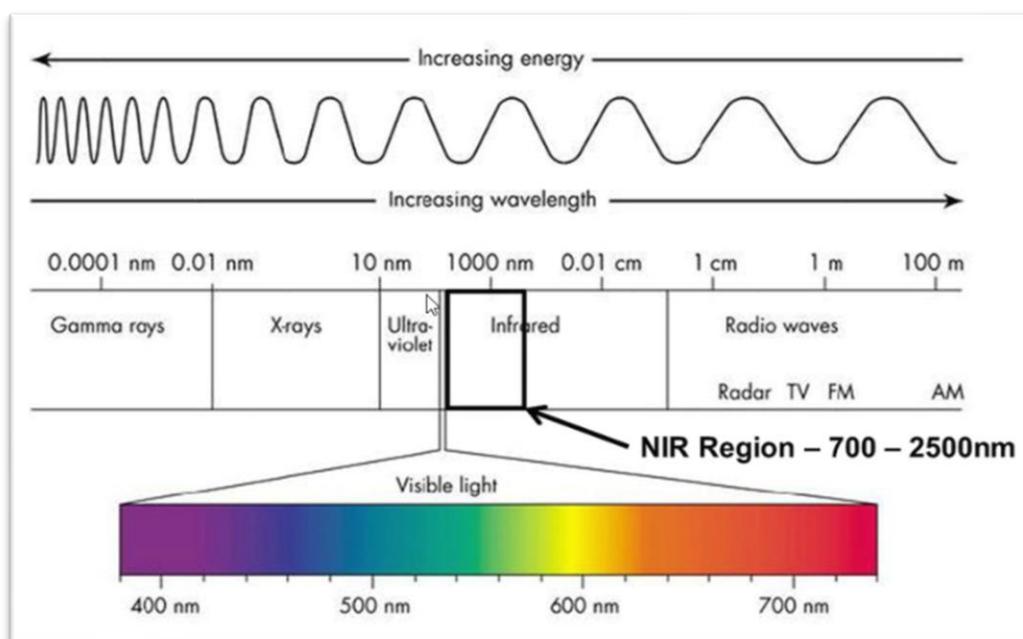


Ilustración 5. Espectro electromagnético

Los usos más importantes de la radiación en el infrarrojo cercano son los relacionados con la determinación cuantitativa de especies, como agua, proteínas, hidrocarburos de bajo peso molecular y grasas en productos agrícolas, alimentos, petróleo y de la industria química. Se utilizan tanto las medidas de reflexión difusa como las de transmisión, aunque la reflectancia difusa es la más utilizada.

Las bandas de los espectros en el infrarrojo cercano son casi siempre anchas y están traslapadas. Rara vez hay bandas espectrales limpias que faciliten la simple correlación

con la concentración del analito, por lo que se requiere aplicar técnicas de calibración de varias variables. Por lo general se usan los mínimos cuadrados parciales, regresión lineal múltiple, regresión de los componentes principales y redes neurales artificiales.

Los fabricantes de instrumentos para el infrarrojo cercano proporcionan programas de computación para crear modelos de calibración. [3]

### 3.4. Quimiometría

#### 3.4.1 Introducción

Las técnicas espectroscópicas generan una gran cantidad de datos relativos a las muestras analizadas. Aunque no quiere decir que obtengamos más información por tener más datos del sistema. Como bien dicen Kowalski y Beebe en un artículo publicado en 1987, los datos solo son valiosos para los químicos y la sociedad cuando son interpretados y utilizables; entonces los datos se convierten en información. [4] Aquí es donde entra la quimiometría, que puede definirse como la disciplina química que utiliza métodos matemáticos, estadísticos y de otro tipo que emplean la lógica formal (a) para diseñar o seleccionar procedimientos de medición y experimentos óptimos, y (b) para proporcionar la máxima información química relevante mediante el análisis de datos químicos. [5]

La quimiometría pretende convertir los datos en información mediante el uso de las siguientes herramientas:

- Pretratamientos para la mejora de la señal instrumental
- Técnicas de calibrado
- Técnicas de clasificación

Dada la complejidad de los espectros generados por el NIR la quimiometría es imprescindible para poder interpretar, entender y modelar el conjunto de datos que este nos entrega. En este apartado presentaremos la base teórica de las herramientas quimiométricas facilitadas por el software. Empezaremos por los métodos de selección de muestra, después los pretratamientos matemáticos para la mejora de la señal

instrumental, las técnicas de calibrado destinadas a la cuantificación y, por último, la validación.

### 3.4.2 Métodos de selección de outliers

En este apartado presentaremos los métodos que utiliza el software para la selección de outliers.

Se pueden utilizar 3 métodos para la selección de muestras: Selección aleatoria, selección por PCA (distancia de Mahalanobis) y selección por distancia de longitud de onda.

#### **Selección aleatoria.**

El método por selección aleatoria selecciona muestras al azar para los conjuntos de calibración y validación. No se crea un conjunto de outliers cuando se usa este método. (El tratamiento matemático no es posible cuando se elige este método)

#### **Selección por distancia de Mahalanobis.**

La selección por PCA primero calcula un modelo de componentes principales para un producto. Todas las muestras con distancias Mahalanobis desde el centro de la distribución superiores a un umbral definido por el usuario se marcan como valores atípicos. Las muestras situadas en regiones de alta densidad de la población (vecinos más cercanos) se identifican como redundantes, de modo que las distancias euclidianas entre las muestras del conjunto de entrenamiento son mayores que el umbral.

#### **Selección por distancia de longitud de onda.**

Las muestras con una distancia máxima del espectro medio del producto superior al umbral se colocan en el conjunto de valores atípicos. La selección de muestras redundantes usa un método similar al de PCA's, excepto que la distancia euclídea se usa en la longitud de onda.

### 3.4.3 Métodos de pretratamiento para la mejora de la señal instrumental

Los componentes o efectos no deseados presentes en las señales obtenidas en espectroscopia se denominan comúnmente ruido. Este ruido puede tener diferentes causas u orígenes, por lo que puede afectar de forma distinta al espectro. Puede haber ruido debido a los componentes de la instrumentación utilizada para el registro del espectro (ruido instrumental), debido a variaciones de temperatura, humedad u otras condiciones ambientales durante el registro (ruido ambiental) o bien variaciones en la señal debidas a la propia naturaleza de la muestra. Para corregir o reducir este ruido de los datos espectrales es necesario un pretratamiento espectral.

A continuación, se comentan los tratamientos espectrales más habituales en espectroscopia NIR, aplicados con el fin de reducir o eliminar la contribución del ruido.

#### **Derivación**

Es uno de los pretratamientos más utilizados en espectroscopia para minimizar desviaciones de línea base causados por efectos de dispersión. En espectroscopia NIR se utiliza para disminuir los problemas más característicos: solapamiento de bandas y variaciones de línea base. La primera derivada elimina los términos constantes a todas las longitudes de onda, es decir, desplazamientos de línea base, mientras que la segunda derivada corrige además las desviaciones causadas por los términos que varían linealmente con la longitud de onda. Generalmente no se utilizan derivadas de orden superior. La aplicación de las derivadas permite un aumento de la resolución de bandas, pero por el contrario aumenta el ruido.

#### **Corrección de la línea base**

Es un tipo de pretratamiento que intenta corregir determinadas tendencias en la línea base que aportan ruido a la señal. Existen varios tipos de corrección de línea base según el efecto que se desea conseguir. Un tipo de corrección es el que modela la línea base como una función simple de longitud de onda y sustrae esta función a todos los datos espectrales.

Cuando se trabaja con Transformada de Fourier, es frecuente utilizar una corrección de la línea base llamada "Tilting" que consiste en corregir el espectro eliminando la curvatura de manera que se cumplan los requisitos de periodicidad para la transformación.

Un último tipo de corrección de línea base es el llamado Ajuste de Línea Base que ajusta el espectro original a una función cuadrática, sustrayéndola posteriormente del espectro original, eliminando la característica curvatura ascendente a medida que aumenta la longitud de onda. De esta forma, el conjunto de espectros corregidos tiene como media el valor 0 y una varianza diferente de cero. Este pretratamiento se aplica a cada espectro en individual, por lo tanto, no depende de ningún espectro de referencia.

### **Corrección del efecto multiplicativo de la dispersión (MSC)**

Este método tiene la finalidad de corregir desplazamientos espectrales debidos al diferente tamaño de partícula de las muestras. Es por lo tanto un método muy utilizado en NIR. La corrección se lleva a cabo aplicando a cada espectro original los parámetros de la recta obtenida por mínimos cuadrados representando los valores de absorbancia de un intervalo de longitudes de onda en el que el analito no absorbe frente a los valores de absorbancia de un espectro de referencia (normalmente el espectro medio del conjunto a tratar). Su principal inconveniente es que, si el conjunto de espectros originales es modificado, el espectro de referencia cambia, por lo que los parámetros de la recta obtenida deben recalcularse de nuevo.

### **Variable Normal Estándar (SNV)**

Es otro pretratamiento utilizado para la corrección de los efectos de tamaño de partícula en espectroscopía NIR. A diferencia de MSC, SNV se aplica individualmente a cada espectro, sin depender de ningún espectro de referencia. De esta forma se obtiene una escala común a todos los espectros. El procedimiento de corrección consiste en centrar cada espectro, restando el valor medio de absorbancia del espectro a cada uno de los valores de absorbancia a cada longitud de onda, dividiendo el resultado por la desviación estándar del espectro. Se ha demostrado que existe una relación lineal entre MSC y SNV, por lo que, aplicados correctamente, proporcionan resultados equivalentes.

Hasta aquí se han revisado los pretratamientos espectrales más comunes en NIR. En ocasiones resulta indicado utilizar más de un método de pretratamiento espectral para mejorar la calidad de los resultados. Sin embargo, esto puede conllevar también una eliminación de información útil, por lo que en cada caso se deberá evaluar qué tratamiento es mejor o bien si es necesario aplicar más de uno para la corrección de datos.

#### **3.4.4 Técnicas de calibrado**

En este apartado hablaremos de las técnicas utilizadas por el software para la calibración de modelos multivariantes.

##### **Regresión múltiple lineal [6]**

Es un método de análisis cuantitativo que requiere de una selección previa de las variables significativas a partir de las cuales se construye la ecuación de regresión. La regresión lineal múltiple establece una relación lineal entre la señal y la propiedad a determinar mediante un ajuste de mínimos cuadrados. La concentración se calcula multiplicando los valores de absorbancia por unos coeficientes de proporcionalidad obtenidos durante la etapa de calibrado.

El desarrollo de modelos MLR requiere una adecuada selección de las longitudes de onda que son utilizadas para su cálculo. Dicha tarea es compleja ya que en muchas ocasiones la variación de la concentración a determinar no puede ser asignada a unas longitudes de onda discretas y es por ello que su uso en espectroscopia NIR es reducido.

##### **Regresión parcial sobre mínimos cuadrados (PLS)**

La PLS empieza encontrando combinaciones lineales de las variables predictoras (factores o componentes principales). 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. Se suele distinguir entre la situación cuando la respuesta consta de una única variable y aquella cuando la respuesta es

multivariante: la primera se llama PLS1, la última PLS2. [7] Normalmente si tenemos (j) variables es mejor hacer j modelos PLS1 que un único modelo PLS2, ya que, cada modelo PLS1 puede ajustarse al número de factores necesarios individualmente.

### 3.4.5 Validación de la recta

Es imprescindible que el modelo demuestre su habilidad cuando se tienen que analizar nuevos objetos no utilizados en la construcción del modelo. Es decir, hace falta validar el modelo y asegurarse que se comportará adecuadamente en el momento de predecir muestras nuevas. Hay dos tipos de técnicas de validación:

#### **Conjunto de entrenamiento y prueba (training set y test set)**

En esta técnica utilizaremos un conjunto de datos para hacer el modelo y otro distinto para validarlo. El conjunto de prueba tiene que ser representativo de las condiciones futuras de aplicación del modelo. Idealmente tiene que haber un número parecido de muestras en los dos conjuntos (como mínimo 20-40% del conjunto total). Es la mejor estrategia de validación, pero se necesitan muchas muestras lo que significa tiempo y dinero.

La comparación con los resultados obtenidos y la concentración real de las muestras se utiliza para calcular el RMSEP (Root Mean Square Error of Prediction). Este parámetro estadístico lo utilizaremos para conocer el error de predicción de nuestro modelo. Como podemos esperar, cuanto menor sea el valor de RMSEP mejor será el modelo.

#### **Validación cruzada**

Es la mejor alternativa a la técnica de validación comentada anteriormente. Este tipo de validación se lleva a cabo extrayendo muestras individuales del conjunto de calibración. Utilizando las muestras restantes se crea un modelo quimiométrico y se utiliza para analizar las muestras previamente separadas. Extrayendo las muestras antes de crear el modelo quimiométrico se garantiza que la muestra sea desconocida e independiente del modelo de calibración.

La acción comentada se va realizando con todas las muestras hasta haber analizado todas las muestras.

La comparación con los resultados obtenidos y la concentración real de las muestras se utiliza para calcular el RMSECV (Root Mean Square Error of Cross Validation). Este parámetro estadístico lo utilizaremos para conocer el error de predicción de nuestro modelo. Como podemos esperar, cuanto menor sea el valor de RMSECV mejor será el modelo.

Para la validación y comparación de los modelos de calibrado utilizaremos también los siguientes parámetros estadísticos:

- BIAS que nos informa del error sistemático del modelo.
- SEP precisión asociada al modelo.
- Linealidad

### 3.5 Ventajas de la utilización del NIR [8]

Las técnicas tradicionales de análisis químico (cromatografía, valoraciones, etc.), son ampliamente utilizadas en la industria química para el control de reacciones. Estas técnicas implican un retraso en la obtención de la información debido al tiempo necesario para realizar las diferentes etapas de la determinación tradicional, como el muestreo, el análisis y el cálculo para proporcionar un resultado.

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

A pesar de que el desarrollo de modelos de calibración es complejo, una vez instalada la calibración, el análisis de rutina puede ser realizado por personal no cualificado.

## 4. Parte experimental

### 4.1 Diagramas de Gantt PVOH escamas y PVOH producto final

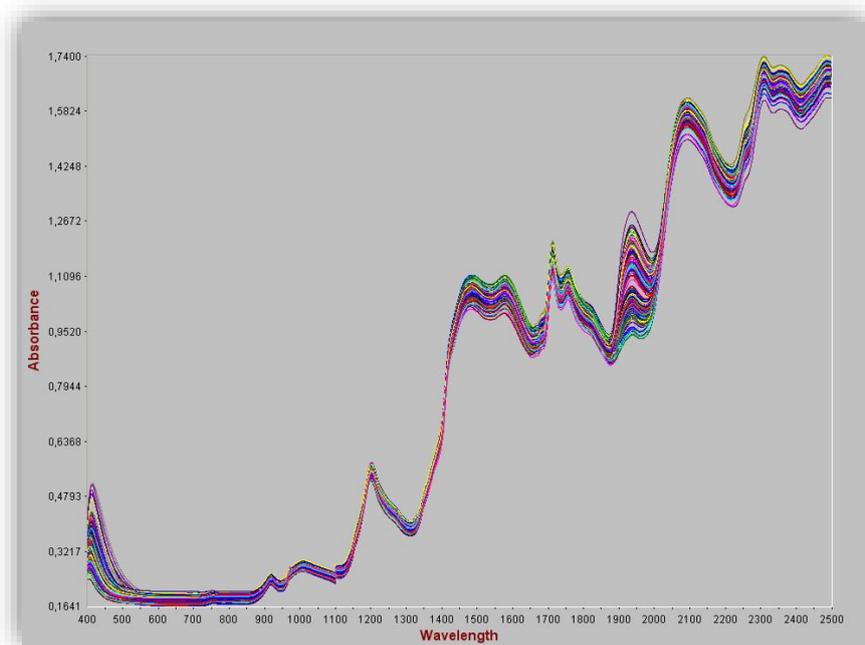
| Tareas         | Estado | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2  | 2 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |   |   |   |
|----------------|--------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|                |        | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 8 | 9 | 0 | 1 | 2 | 5 | 6 | 7 | 8 | 9 | 2 | 3  | 2 | 2 | 2 | 3 | 0 | 1 | 1 | 2 | 5 | 6 | 7 | 8 | 9 | 0 | 2 | 3 | 4 | 5 |
|                |        | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 4- | 2 | 2 | 2 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
|                |        | f | f | f | f | m | m | m | m | m | m | m | m | m | m | m | m | m | m | m | m | r  | m | m | m | m | m | a | a | a | a | a | a | a | a | a | a | a | a |   |
|                |        | e | e | e | e | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a  | a | a | a | a | b | b | b | b | b | b | b | b | b | b | b | b |   |   |
|                |        | b | b | b | b | r | r | r | r | r | r | r | r | r | r | r | r | r | r | r | r | r  | r | r | r | r | r | r | r | r | r | r | r | r | r | r | r | r | r |   |
| Inicio         |        |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Form. interna  | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Form. seguir.  | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Bibliografía   | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Form. NIR      | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Consulta prof. | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Desarrollo     |        |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Obt. espec.    | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Desar. Modelo  | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Final          |        |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Val. modelo    | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Redactado      | OK     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

| Tareas         | Estado | 20-abr | 21-abr | 22-abr | 23-abr | 26-abr | 27-abr | 28-abr | 29-abr | 30-abr | 03-may | 04-may | 05-may | 06-may | 07-may | 10-may | 11-may | 12-may | 13-may | 14-may | 17-may | 18-may | 19-may | 20-may | 21-may | 24-may | 25-may | 26-may | 27-may | 28-may |  |
|----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--|
| Desarrollo     |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |  |
| Obt. de espec. | OK     |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |  |
| Desar. Modelo  | OK     |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |  |
| Final          |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |  |
| Val. modelo    | OK     |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |  |
| Redactado      | OK     |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |  |

## 4.2 Adquisición de espectros para el modelo multivariante

Todo que en este laboratorio ya existe un modelo de calibración para este rango de grado de hidrólisis se ha visto que no funciona bien para este producto, por ese motivo procedemos a pasar nuevas muestras de distintas campañas. Las muestras se han analizado de la siguiente manera:

- Parámetros de filtro espectral: reflectancia con un rango de 400 a 2500 nm.
- Número de escaneos: 32
- Análisis realizado por triplicado
- Número de muestras analizado: 197



*Ilustración 6. Espectros de las 197 muestras.*

En este caso no podemos agrandar el rango de la recta de calibración ya que no podemos aumentar o disminuir el GH a nuestro antojo. La especificación del producto tiene un rango de GH de 86 a 90% y las muestras de la calibración de 85,82 a 90,40.

Después de analizar las muestras por triplicado, haremos otro conjunto de espectros promediando los 3 análisis que hemos hecho por muestra.

### 4.3 Selección de muestras

Una vez tenemos el conjunto de muestras procederemos a realizar la selección de las mismas, en este caso para diferenciarlas utilizaremos la distancia de Mahalanobis y la distancia de longitud de onda.

El objetivo de la selección de muestras es identificar los valores atípicos y las muestras redundantes. Durante la selección de muestras, los sospechosos de ser valores atípicos y los espectros redundantes se identifican según el umbral que el usuario puede modificar. Los valores atípicos se colocan en el conjunto de rechazo y las muestras redundantes en el conjunto de aceptación. Previamente se habrá seleccionado que tipo de pretratamiento para la mejora de la señal instrumental aplicar y el porcentaje de muestras a utilizar para crear el conjunto de validación (en este caso 20%) como podemos ver en las ilustraciones 7 y 8.

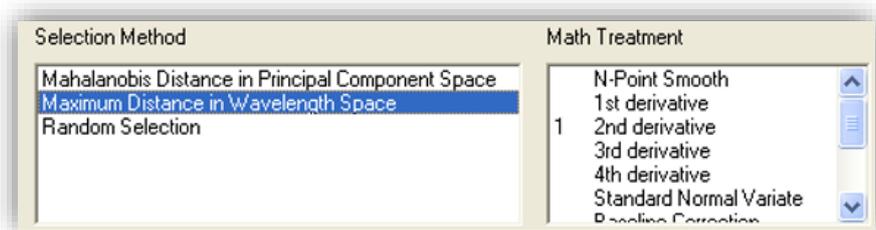


Ilustración 7. Ventana de la interfaz del software la cual nos deja elegir el método de selección de muestra y el pretratamiento.

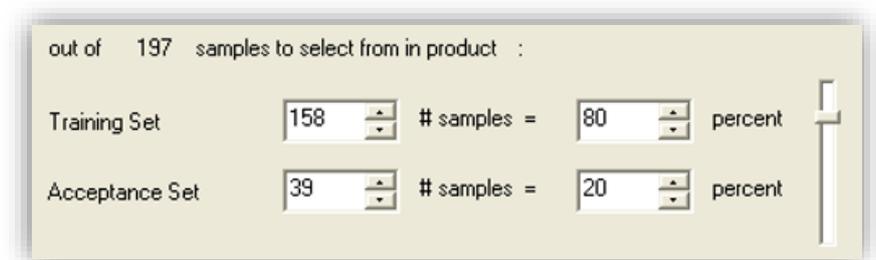
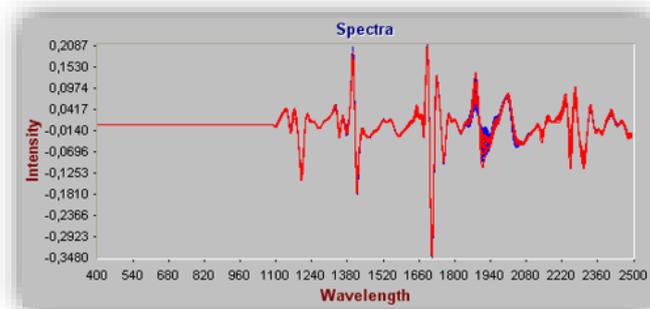


Ilustración 8. Ventana de la interfaz del software la cual nos deja elegir la proporción de muestras que irán a cada conjunto.

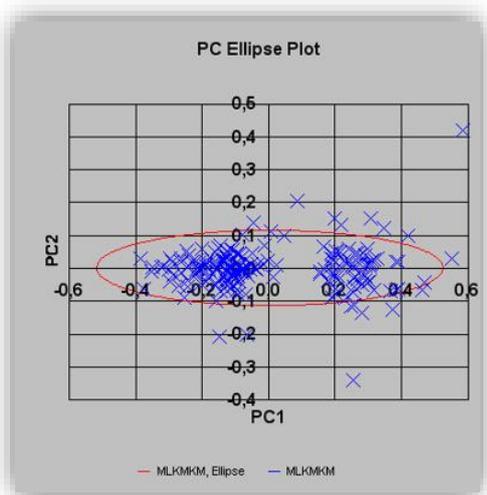
**Distancia de Mahalanobis**

En las siguientes imágenes se muestra la información proporcionada por la técnica de selección de muestra distancia de Mahalanobis.

En la ilustración 9 podemos ver todos los espectros agrupados, marcándose en rojo los considerados outliers.

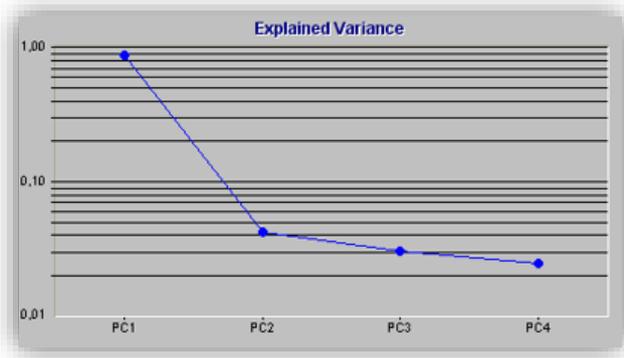


*Ilustración 9. Todos los espectros con el pretratamiento donde se marca en rojo las muestras seleccionadas por outliers*



*Ilustración 10. Umbral de restricción del método de selección de muestras por distancia de Mahalanobis*

El umbral del método define los límites del elipsoide del modelo. Durante el análisis, las muestras que se encuentren fuera del elipsoide no serán identificadas ni calificadas. En la ilustración 10 podemos ver el umbral marcado en rojo.



En la ilustración 11 podemos ver la varianza explicada por cada PC

Ilustración 11. Gráfica que representa la varianza explicada por cada PC

### Distancia de longitud de onda

En las siguientes imágenes podemos ver como la técnica de selección de muestra está detectando outliers. En este caso en concreto pusimos que las muestras a distancia superior de 5 sean reconocidas como outlier. En la ilustración 12 podemos ver los espectros de las muestras agrupadas, estando marcadas en rojo las muestras consideradas como outliers.

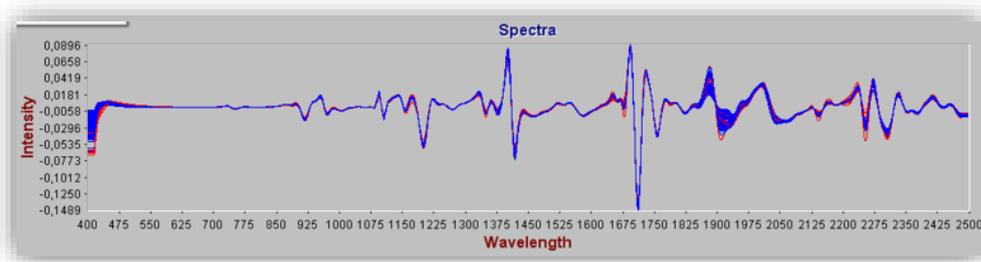


Ilustración 12. Todos los espectros con el pretratamiento donde se marca en rojo las muestras seleccionadas por outliers.

En la ilustración 13 vemos la distancia que hay entre las muestras y la mediana de los espectros. Los más alejados de 5 son marcados en rojo como outliers.

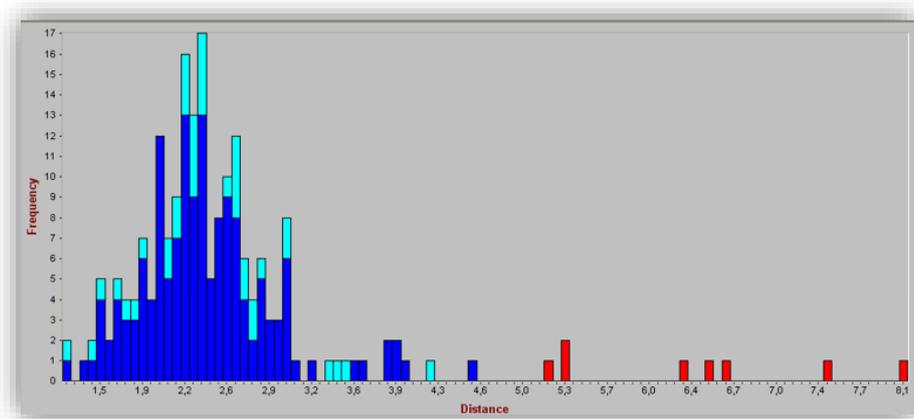


Ilustración 13. Representación de la distancia y la frecuencia marcando en rojo los outliers.

En nuestro caso vamos a probar los dos métodos para ver cual nos terminará dando un mejor modelo.

#### 4.4 Desarrollo de los modelos de calibrado multivariante

El software nos hará elegir uno o varios pretratamientos matemáticos para la mejora de la señal instrumental y la técnica de calibrado a utilizar. En nuestro caso utilizaremos la técnica de calibración PLS (Regresión parcial sobre mínimos cuadrados) y como pretratamiento probaremos con la segunda derivada únicamente y con segunda derivada más SNV (debido a que las escamas tienen un tamaño de partícula muy irregular). También debemos marcarle el número máximo de factores a calcular y el rango de longitud de onda en el que vamos a realizar la calibración.

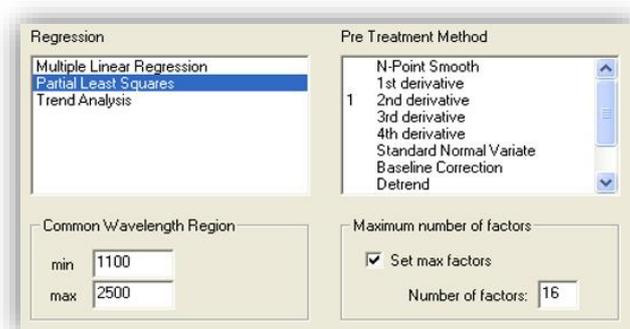


Ilustración 14 Ventana de la interfaz del software la cual nos deja elegir el método de calibración y el pretratamiento

Al haber elegido PLS el siguiente paso será elegir la validación que vamos a utilizar, nosotros usaremos validación cruzada dejando fuera 4 muestras cada vez, pero como podemos ver en la ilustración 15 tenemos varias opciones para realizar la validación.

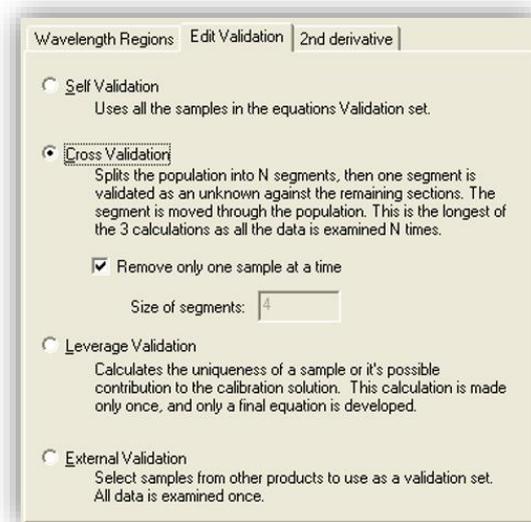


Ilustración 15. Ventana de la interfaz del software la cual nos deja elegir el tipo de validación interna que vamos a realizar.

Una vez completados los pasos mencionados anteriormente obtendremos información sobre el modelo.

En el gráfico de arriba a la izquierda llamado “Loadings” (ilustración 16) podemos ver el valor de los coeficientes para cada variable. Hay un gráfico de Loadings diferente para cada factor. Cada factor expresa un % de la información explicada, todos los PCs juntos expresan el 100% de la información (la útil y el ruido).

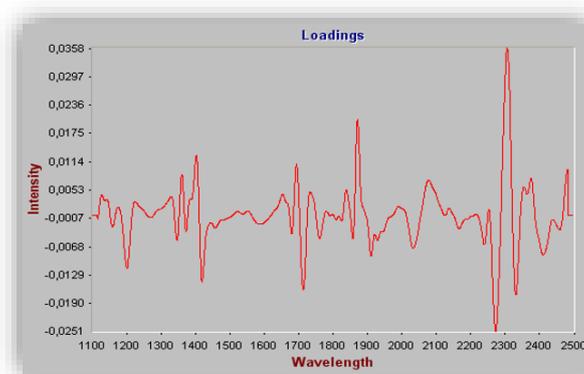


Ilustración 16. Gráfico de Loadings

La siguiente pestaña, arriba a la derecha, nos muestra el gráfico que nos indica el calor de PRESS en función del número de factores utilizados (ilustración 17). El software nos elegirá como factor el que tenga un valor menor de PRESS. El estadístico PRESS (suma de cuadrados del error residual predicho) es una medida de la validación cruzada en el análisis de regresión.

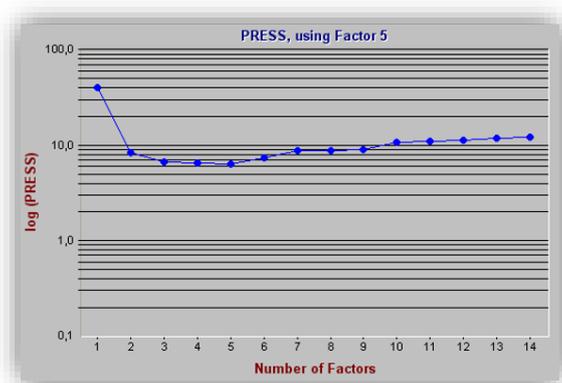


Ilustración 17. Representación gráfica del PRESS para cada factor

En el gráfico de abajo a la izquierda llamado Calibración Set: Calculated vs Lab Data (ilustración 18) tenemos las muestras distribuidas por el valor de GH dado por el NIR vs el del análisis de referencia. Vemos una línea diagonal que marca la idealidad, es decir, un mismo valor para el resultado del análisis de referencia y la predicción del modelo. Con esto, de manera visual podremos saber cuánto se acercan o se alejan de la idealidad cada una de las muestras predichas.

Dentro de esta pestaña tenemos varios apartados en el que podemos ver también un resumen de los valores del modelo (ilustración 20) y un gráfico de residuales (ilustración 14) para poder ver la distancia de las muestras a la idealidad de una forma todavía más visual.

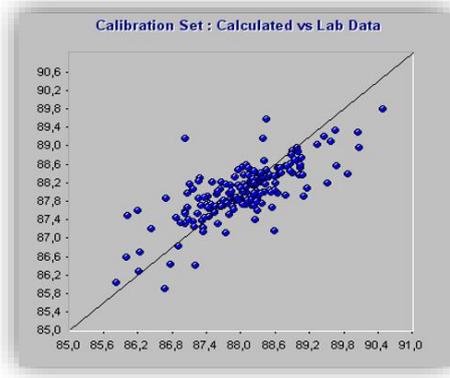


Ilustración 18. Gráfico Calibration Set: Calculated vs Lab Data

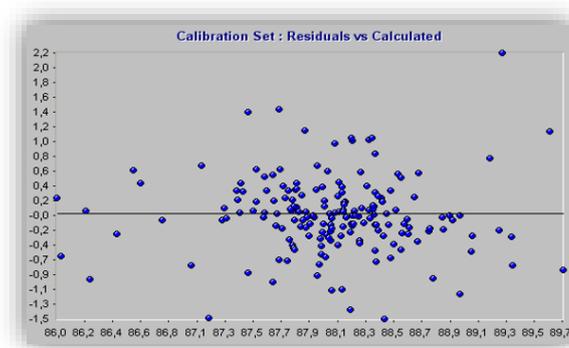


Ilustración 19. Gráfico Calibration Set: Residuals vs Calculated

| Factor                   | R Squared | Error Cal | PRESS   | F-Value  | dErr X Validati |
|--------------------------|-----------|-----------|---------|----------|-----------------|
| ---> Equation Constants: |           |           |         |          |                 |
| Factor 1                 | 0,7981    | 0,3235    | 39,9013 | 490,1359 | 0,5627          |
| Factor 2                 | 0,8863    | 0,2437    | 8,3072  | 479,2351 | 0,2568          |
| Factor 3                 | 0,9088    | 0,2192    | 6,6623  | 405,0962 | 0,2299          |
| Factor 4                 | 0,9149    | 0,2126    | 6,4849  | 325,25   | 0,2269          |
| Factor 5 *               | 0,9162    | 0,2119    | 6,3918  | 262,2558 | 0,2252          |
| Factor 6                 | 0,9192    | 0,2089    | 7,4422  | 225,5084 | 0,243           |
| Factor 7                 | 0,9247    | 0,2025    | 8,7164  | 206,9681 | 0,263           |
| Factor 8                 | 0,9273    | 0,1998    | 8,8939  | 186,5659 | 0,2657          |
| Factor 9                 | 0,9318    | 0,1943    | 8,9909  | 176,2327 | 0,2671          |
| Factor 10                | 0,9392    | 0,1843    | 10,5993 | 177,6117 | 0,29            |
| Factor 11                | 0,9449    | 0,1762    | 10,9274 | 177,7031 | 0,2945          |
| Factor 12                | 0,9489    | 0,1704    | 11,2764 | 174,8752 | 0,2992          |
| Factor 13                | 0,9523    | 0,1654    | 11,7417 | 172,152  | 0,3053          |
| Factor 14                | 0,9555    | 0,1605    | 12,0759 | 170,1703 | 0,3096          |
| Factor 15                | 0,9574    | 0,1578    | 12,8157 | 164,6801 | 0,3189          |
| Factor 16                | 0,9605    | 0,1525    | 13,9214 | 165,7725 | 0,3324          |

Ilustración 20. Tabla resumen para cada factor

Desde el gráfico de Calibration Set: Calculated vs Lab Data podemos seleccionar muestras y enviarlas al conjunto de outliers. En el procedimiento que seguimos

marcamos como límite de residual 0.5%. Pensamos que es un valor adecuado ya que el análisis de referencia tiene una precisión de 0.14%.

Una vez terminada la selección de outliers pasaremos a guardar el modelo y realizar la validación externa.

#### 4.5 Validación del modelo de calibrado multivariante

Ahora procederemos a predecir el conjunto de validación con el modelo de calibrado. Para ello debemos ir al apartado de predicción, selecciona la ecuación del modelo, el constituyente de interés y el conjunto a predecir. Una vez hecho esto obtendremos la gráfica Validation Set: Calculated vs Lab Data que tiene la misma función que la Calibration Set explicada anteriormente, pero sin posibilidad, como es normal, de enviar muestras al conjunto de outliers. También obtendremos la gráfica Calculated vs Lab Data pero esta vez del conjunto de validación (ilustración 22), un resumen donde obtendremos la información que podemos ver en la ilustración 21. Y por último una hoja de cálculo con el valor del método de referencia, el predicho y el residual (ilustración 23).

| Term                                      | Value   | Type | X-Value | E | F | G |
|---|---------|------|---------|---|---|---|
| ---> Summary:                             |         |      |         |   |   |   |
| SEP before                                | 0,1773  |      |         |   |   |   |
| ---> Suggested Bias Only Correction:      |         |      |         |   |   |   |
| Bias Adjus                                | 0,0539  |      |         |   |   |   |
| ---> Suggested Slope and Bias Correction: |         |      |         |   |   |   |
| Correlation                               | 0,9509  |      |         |   |   |   |
| Slope Adju                                | 1,0233  |      |         |   |   |   |
| Slope Adj                                 | 0,0571  |      |         |   |   |   |
| Bias Adjus                                | -2,0013 |      |         |   |   |   |
| Bias Adj S                                | 5,0336  |      |         |   |   |   |
| SEP possi                                 | 0,1684  |      |         |   |   |   |

*Ilustración 21. Tabla resumen de la validación externa*

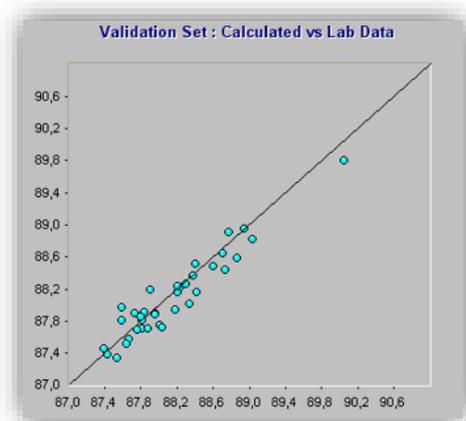


Ilustración 22. Gráfico Validation Set: Calculated vs Lab Data

| Set        | Product Name               | Sample Name | Computed % | Lab % | Residual | Gr |
|------------|----------------------------|-------------|------------|-------|----------|----|
| Validation | PVOH_575_ES(E575 G355 07:2 |             | 87,90      | 87,73 | 0,17     |    |
| Validation | PVOH_575_ES(E575 G355 11:0 |             | 87,71      | 87,81 | -0,10    |    |
| Validation | PVOH_575_ES(E575 G355 23:0 |             | 87,92      | 87,83 | 0,09     |    |
| Validation | PVOH_575_ES(E575 G355 07:0 |             | 88,37      | 88,38 | -0,01    |    |
| Validation | PVOH_575_ES(E575 G355 07:0 |             | 87,58      | 87,66 | -0,08    |    |
| Validation | PVOH_575_ES(E575 G355 03:1 |             | 87,52      | 87,64 | -0,12    |    |
| Validation | PVOH_575_ES(E575 G355 15:0 |             | 88,45      | 88,73 | -0,28    |    |
| Validation | PVOH_575_ES(E575 G355 15:1 |             | 88,17      | 88,42 | -0,25    |    |
| Validation | PVOH_575_ES(E575 G355 03:0 |             | 88,22      | 88,23 | -0,01    |    |
| Validation | PVOH_575_ES(E575 G355 23:0 |             | 87,91      | 87,96 | -0,05    |    |
| Validation | PVOH_575_ES(E575 G355 10:0 |             | 89,80      | 90,05 | -0,25    |    |
| Validation | PVOH_575_ES(E575 G355 02:4 |             | 87,69      | 87,76 | -0,07    |    |
| Validation | PVOH_575_ES(E575 G355 11:0 |             | 87,34      | 87,53 | -0,19    |    |
| Validation | PVOH_575_ES(E575 G355 11:1 |             | 88,02      | 88,34 | -0,32    |    |
| Validation | PVOH_575_ES(E575 G355 07:0 |             | 87,39      | 87,43 | -0,04    |    |
| Validation | PVOH_575_ES(E575 G355 15:0 |             | 88,26      | 88,30 | -0,04    |    |
| Validation | PVOH_575_ES(E575 G355 15:1 |             | 88,16      | 88,21 | -0,05    |    |
| Validation | PVOH_575_ES(E575 G355 23:1 |             | 88,59      | 88,86 | -0,27    |    |

Ilustración 23. Tabla de residuales

El software nos da la posibilidad de exportar la hoja de cálculo nombrada al final del párrafo anterior. Con esos resultados podremos calcular el siguiente conjunto de parámetros estadísticos que nos ayudará a decidir cuál es el mejor modelo.

- RMSECV, expresa el error medio esperable asociado de muestras futuras, calculado a partir de la validación cruzada. Se calcula de la siguiente manera:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i^{NIR} - y_i^{Ref})^2}{n}}$$

Ecuación 1. RMSECV

Donde:

$y_i^{NIR}$  = valor predicho por NIR

$y_i^{Ref}$  = valor dado por el método de referencia

$n$  = Número de muestras

- RMSEP, expresa el error medio esperable asociado de muestras futuras, calculado a partir de la validación externa. Se calcula de la siguiente manera:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i^{NIR} - y_i^{Ref})^2}{n}}$$

*Ecuación 2. RMSEP*

Donde:

$y_i^{NIR}$  = valor predicho por NIR

$y_i^{Ref}$  = valor dado por el método de referencia

$n$  = Número de muestras

$$y_{futura} = y_{predicha} \pm 2 \cdot RMSEP$$

*Ecuación 3. Intervalo de confianza*

- BIAS, expresa el error sistemático del modelo

$$BIAS = \frac{\sum_{i=1}^n (y_i^{NIR} - y_i^{Ref})}{n}$$

*Ecuación 4. BIAS*

Donde:

$y_i^{NIR}$  = valor predicho por NIR

$y_i^{Ref}$  = valor dado por el método de referencia

$n$  = Número de muestras

- SEP, expresa la precisión asociada al modelo

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i^{NIR} - y_i^{Ref} - bias)^2}{n-1}}$$

Ecuación 5. SEP

Donde:

$y_i^{NIR}$  = valor predicho por NIR

$y_i^{Ref}$  = valor dado por el método de referencia

$n$  = Número de muestras

Bias = BIAS calculado anteriormente

- Linealidad ( $R^2$ )

## 5. Resultados y discusión

En este apartado vamos a explicar los resultados obtenidos después de seguir el procedimiento descrito en el apartado anterior, hemos creado 4 modelos de calibrado y compararemos el mejor de ellos con el modelo instaurado en el laboratorio para la determinación del grado de hidrólisis.

También se mostrarán los resultados del modelo para la determinación del grado de hidrólisis para el producto final.

**Comparación de los cuatro modelos para escamas**

Para la creación de los 4 modelos se ha utilizado la técnica de calibrado PLS, la diferencia entre ellos es el método de selección de muestra y el pretratamiento matemático para la mejora de la señal instrumental.

En un principio el objetivo era hacer un modelo calcado al ya instaurado en el laboratorio, pero utilizando muestras de un único producto para hacer la calibración. Pero al empezar a hacer la calibración nos dimos cuenta de que había una gran cantidad de outliers, así que decidimos probar alternativas para ver si mejoraba.

Siguiendo el criterio de detección de outliers comentado anteriormente (residual superior a 0,5%) el resultado de los parámetros estadísticos de cada uno de los modelos son los siguientes:

Tabla 1. Resumen de los resultados de la validación interna de los calibrados multivariantes realizados para escamas.

| Validación interna (GH) |        |       |        |          |                       |                |              |          |
|-------------------------|--------|-------|--------|----------|-----------------------|----------------|--------------|----------|
| R <sup>2</sup>          | RMSECV | SEP   | Bias   | Factores | Selección de muestras | Téc. calibrado | Pretrat.     | Outliers |
| 0,958                   | 0,200  | 0,200 | -0,001 | 5        | MD                    | PLS            | 2ª Der.+SNV  | 41       |
| 0,960                   | 0,200  | 0,200 | 0,003  | 5        | Longitud de onda      | PLS            | 2ª Der.      | 42       |
| 0,964                   | 0,188  | 0,188 | 0,000  | 5        | MD                    | PLS            | 2ª Der.      | 43       |
| 0,955                   | 0,210  | 0,210 | 0,000  | 5        | Longitud de onda      | PLS            | 2ª Der. +SNV | 40       |

Tabla 2. Resumen de los resultados de la validación externa de los calibrados multivariantes realizados para escamas.

| Validación externa (GH) |       |       |        |          |                       |                |             |              |
|-------------------------|-------|-------|--------|----------|-----------------------|----------------|-------------|--------------|
| R <sup>2</sup>          | RMSEP | SEP   | Bias   | Factores | Selección de muestras | Téc. calibrado | Pretrat.    | Núm muestras |
| 0,858                   | 0,197 | 0,196 | -0,020 | 5        | MD                    | PLS            | 2ª Der. SNV | 27           |
| 0,916                   | 0,205 | 0,203 | 0,028  | 5        | Longitud de onda      | PLS            | 2ª Der.     | 29           |
| 0,904                   | 0,232 | 0,216 | 0,086  | 5        | MD                    | PLS            | 2ª Der.     | 27           |
| 0,904                   | 0,177 | 0,169 | -0,054 | 5        | Longitud de onda      | PLS            | 2ª Der.+SNV | 36           |

En las tablas 1 y 2 tenemos los resultados de los tipos de validaciones, como podemos ver, los resultados tanto para la validación interna como para la externa para los diferentes modelos no son muy diferentes. Ya que el método de validación más fiable es

el de validación externa (ya que tienes dos conjuntos independientes) y dentro de este el que parece más robusto es el modelo hecho a partir de una selección de muestra por distancia de longitud de onda y pretratamiento de segunda derivada más SNV (menor valor de RMSEP (es decir intervalo de confianza menor), de SEP y BIAS), será este el que elijamos para comprar con el método ya establecido en el laboratorio.

### **Comparación con el modelo del laboratorio de escamas**

Para la comparación del modelo de laboratorio calculamos los mismos valores estadísticos que los demás modelos, utilizando el conjunto de validación del modelo elegido. También vale la pena añadir que la selección de muestra y pretratamientos espectrales son diferentes. Habiendo sido utilizado para el modelo antiguo del laboratorio, distancia de longitud de onda como método de selección de muestra y como pretratamiento únicamente segunda derivada. Los resultados son los siguientes:

Tabla 3. Parámetros estadísticos del modelo seleccionado y del modelo utilizado en el laboratorio

| Validación externa (GH) |       |       |        |                    |
|-------------------------|-------|-------|--------|--------------------|
| R <sup>2</sup>          | RMSEP | SEP   | Bias   | Número de muestras |
| 0,881                   | 0,400 | 0,209 | -0,341 | 36                 |
| 0,904                   | 0,177 | 0,169 | -0,054 | 36                 |

Podemos ver en la tabla 3 que para el producto de interés el nuevo modelo (segunda fila) se ajusta mucho más, obteniendo valores más apropiados en todos los parámetros estadísticos. En las ilustraciones 24 y 25 vemos la recta de la predicción del conjunto de validación para cada uno de los modelos.

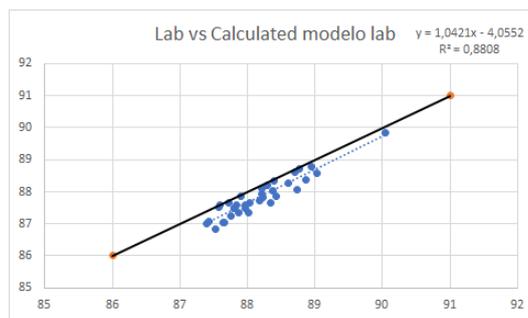


Ilustración 24. Gráfica Lab vs Calculated del conjunto de validación predicho por el modelo del laboratorio

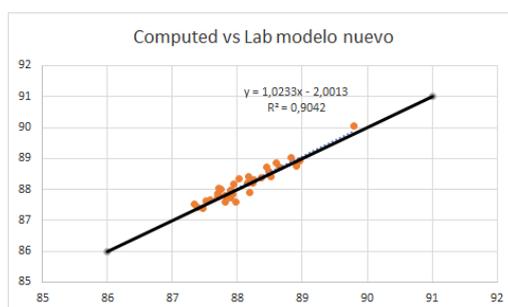


Ilustración 25. Gráfica Lab vs Calculated del conjunto de validación predicho por el modelo nuevo

En las ilustraciones 24 y 25 se muestra de manera muy visual como el modelo anterior predice para todo el rango de la calibración, por debajo de la concentración real, mientras que el modelo nuevo se ajusta mucho más al valor real de la muestra.

Los intervalos de confianza para cada uno de los modelos son los siguientes:

Modelo nuevo =  $\pm 0.35\%$

Modelo lab =  $\pm 0.80\%$

### Resultados del modelo para producto final

En este caso hemos utilizado la experiencia de los modelos hechos para escamas para optimizar el proceso de selección de método de selección de muestra y pretratamiento matemático para la mejora de la señal instrumental.

Se han realizado 3 modelos de calibración distintos y los resultados de los parámetros estadísticos de cada uno de ellos son los siguientes:

Tabla 4. Resumen de los resultados de la validación interna de los calibrados multivariantes realizados para producto final.

| Validación interna (GH) |        |       |          |        |                       |                |                |               |          |
|-------------------------|--------|-------|----------|--------|-----------------------|----------------|----------------|---------------|----------|
| R <sup>2</sup>          | RMSECV | SEP   | Bias     | Factor | Selección de muestras | Téc. calibrado | Pretratamiento | Núm. muestras | Outliers |
| 0,848                   | 0,166  | 0,166 | -7,7E-06 | 5      | Longitud de onda      | PLS            | 2ª Der. +SNV   | 59            | 0        |
| 0,851                   | 0,184  | 0,184 | -3,5E-07 | 5      | Longitud de onda      | PLS            | 2ª Der. +SNV   | 43            | 0        |
| 0,813                   | 0,186  | 0,186 | -9,6E-06 | 5      | Longitud de onda      | PLS            | 2ª Der +SNV    | 74            | 0        |

Tabla 5. Resumen de los resultados de la validación externa de los calibrados multivariantes realizados para producto final.

| Validación externa (GH) |       |       |        |        |                       |                |                |               |  |
|-------------------------|-------|-------|--------|--------|-----------------------|----------------|----------------|---------------|--|
| R <sup>2</sup>          | RMSEP | SEP   | Bias   | Factor | Selección de muestras | Téc. calibrado | Pretratamiento | Núm. muestras |  |
| 0,705                   | 0,241 | 0,239 | -0,034 | 5      | Longitud de onda      | PLS            | 2ª Der +SNV    | 15            |  |
| 0,765                   | 0,209 | 0,179 | 0,002  | 5      | Longitud de onda      | PLS            | 2ª Der +SNV    | 11            |  |

En las tablas 4 y 5 tenemos los resultados de los dos tipos de validaciones. Empezamos haciendo la calibración de la segunda fila de la tabla 4, pero pensamos que eran pocas muestras y decidimos pasar más muestras para ver si mejorábamos el modelo de calibración. Seguidamente hicimos dos modelos más, el de la primera fila de la tabla 4 y el de la tercera fila de esta misma tabla. En el primero se crearon dos conjuntos de muestras, el de calibración y el de validación y en otro al no tener un gran número de muestras pensamos en tan solo crear el conjunto de calibración y realizar únicamente la validación interna.

En esta ocasión, nos guiaremos por los resultados de la validación interna ya que, aunque el método de validación externa es más fiable, tenemos un número pequeño de muestras, así que pensamos que el resultado de la validación interna es más representativo.

Como podemos ver en la validación interna los resultados son muy parecidos entre ellos, elegimos el de mejores valores estadísticos, en este caso el de la fila 1 de la tabla 4.

El intervalo de confianza para el modelo de producto final es el siguiente:

Modelo producto final:  $\pm 0.3319\%$

## 6. Conclusión

Del estudio realizado para la comparación de dos modelos multivariantes mediante espectroscopia NIR y preprocesamiento de datos espectrales con técnicas de calibración multivariable, se pueden extraer las siguientes conclusiones:

### Conclusiones generales:

- La espectroscopia NIR ha demostrado una vez más ser una técnica útil para la determinación del grado de hidrólisis en muestras de alcohol polivinílico.
- Los espectros NIR presentan bandas anchas y solapadas, de modo que una identificación directa por comparación es inviable. Cosa que podemos hacer gracias a la quimiometría.
- Pienso que si hubiésemos tenido un rango de grado de hidrólisis más amplio en el conjunto de calibración el modelo hubiese salido mejor. Pero no tenemos la posibilidad en el laboratorio de reducir o aumentar el grado de hidrólisis. Y el rango dado por producción está dentro de especificaciones.

### Conclusiones fase intermedia del proceso (escamas)

- Se han construido distintos modelos quimiométricos para determinar del grado de hidrólisis de las muestras de alcohol polivinílico. Los cuatro perfiles estudiados con distintas técnicas de selección de muestra y pretratamientos para la mejora de la señal instrumental proporcionan resultados similares en cuanto a parámetros estadísticos y demuestran una predicción correcta.
- La conclusión principal de esta parte del trabajo es que al comprar el modelo elegido entre los 4 y el implantado en el laboratorio vemos una mejora considerable en la predicción. Lo que quiere decir que el trabajo realizado ha aportado una mejora al control analítico de este producto.

### Conclusiones producto final

- Se puede concluir que se ha desarrollado un buen modelo para la determinación del grado de hidrólisis para el producto final de un producto de alcohol polivinílico. Con esto se ha aportado una mejora al control analítico del producto.

## 6. Conclusion

From the study carried out for the comparison of two multivariate models using NIR spectroscopy and preprocessing of spectral data with multivariate calibration techniques, the following conclusions can be drawn:

### General conclusions:

- NIR spectroscopy has once again proven to be a useful technique for the determination of the degree of hydrolysis in polyvinyl alcohol samples.
- NIR spectra present broad and overlapping bands, so that a direct identification by comparison is unfeasible. Which we can do thanks to chemometrics.
- I think that if we had had a wider concentration range in the calibration set the model would have come out better. But we don't have the possibility in the lab to reduce or increase the degree of hydrolysis in a controlled way. And the range given by production is within specifications.

### Conclusions intermediate stage of the process (flakes)

- Different chemometric models have been constructed to determine the degree of hydrolysis of polyvinyl alcohol samples. The four profiles studied with different sample selection techniques and pretreatments for instrumental signal enhancement give similar results in terms of statistical parameters and demonstrate a correct prediction.
- The main conclusion of this part of the work is that when comparing the model chosen among the 4 and the one implemented in the laboratory we see a considerable improvement in the prediction. This means that the work carried out has improved the analytical control of this product.

Conclusions final product

- It can be concluded that a good model has been developed for the determination of the degree of hydrolysis for the final product of a polyvinyl alcohol product. With this, an improvement has been brought to the analytical control of the product.

## 7. Bibliografía

- [1] Sivia, Z.; Roberto, M. R.; Natalia, S.; Marcos, S. Espectroscopia por infrarrojo cercano (NIRS). Su aplicación en análisis de jugos de caña de azúcar. Rev. Ind. agric. Tucumán. [Online] 2010, 87, Artículo 1. [http://www.scielo.org.ar/scielo.php?pid=S1851-30182010000100001&script=sci\\_arttext](http://www.scielo.org.ar/scielo.php?pid=S1851-30182010000100001&script=sci_arttext) (accedido el marzo 22, 2021)
- [2] C. A. Finch; *Polyvinyl Alcohol-Developments*; Editorial Wiley: Aylesbury, 1973.
- [3] D. A. Skoog, F. J. Holler, T. A. Nieman; *Principios de análisis instrumental*; Editorial Cengage Learning: Santa Fe, 2008.
- [4] K. R. Beebe y B. R. Kowalski; An introduction to multivariate calibration analysis. *Anal. Chem.* **1987**, 59, 1007A, 1012A, 1014A.
- [5] D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufmann; *Chemometrics: A textbook*. Editorial Elsevier: Amsterdam, 1988.
- [6] Peguero Gutiérrez, Anna. La espectroscopia NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabado. Tesis Doctoral. Programa de doctorado de Química. Departamento de Química. Facultad de Ciencias UAB, mayo de 2010.
- [7] J. C. Miller, J. N. Miller; *Estadística y Quimiometría para Química Analítica*; Editorial Prentice Hall: Madrid, 2002.
- [8] Castillo Martínez, Miguel. Aplicación de la Espectroscopia NIR al control Analítico de Procesos de la Industria. Tesis doctoral. Programa de doctorado de Química. Facultad de Ciencias UAB, mayo de 2007.