



Un Libro Blanco elaborado por FOSS

## EL RINCÓN QUIMIOMÉTRICO

### **Calificación: Análisis de la adulteración con la tecnología NIR – un estudio sobre la leche desnatada en polvo**

Se presentan los métodos cualitativos para el análisis de la adulteración, así como un caso concreto: la detección de melamina en la leche desnatada en polvo mediante la tecnología NIR y la tecnología quimiométrica.

Autores: Lars Nørgaard\*, Mark Westerhaus, Karin Kjeldahl, Ib Haunstrup, FOSS

\*lno@foss.dk



## Introducción

La buena calidad en la producción de alimentos y piensos se basa en la determinación fiable de la composición de las materias primas, así como de los productos intermedios y finales. No obstante, también resulta igual de importante poder detectar espectros de muestras que presenten desviaciones y que puedan estar causados por adulteraciones, percances durante el proceso o desviaciones de la calidad. La detección de muestras que presentan desviaciones frente a lo que es un buen producto se conoce como *cuálificación* – se recomienda consultar el Libro Blanco InFocus sobre terminología para obtener una descripción de los términos *identificación* y *cuálificación* [1].

La necesidad de contar con métodos eficaces para la detección de adulteraciones y otras desviaciones resulta obvia; en este estudio, describimos la base del método y ofrecemos un ejemplo de cómo se pueden aplicar los métodos para la detección de adulteraciones en el análisis de productos lácteos en polvo mediante la solución NIRS DS2500 de FOSS, combinada con la versión más actualizada del software quimiométrico WinISI.

## Las matemáticas

Como base para el desarrollo del modelo cualitativo, contamos con un conjunto de datos de instrucción con un buen espectro de productos procedentes de una serie de muestras que abarcan la variación esperada del producto pertinente. Los métodos aplicados se basan en el Análisis de Componentes Principales (ACP, o PCA en inglés) y las distancias derivadas en un modelo de ACP de los datos. Como se puede apreciar en la Figura 1, el ACP se puede describir como una descomposición de los datos espectrales preprocesados,  $\mathbf{X}$ , en una matriz  $\mathbf{T}$ ,  $\mathbf{P}$  y  $\mathbf{E}$  para los componentes  $A$ . La matriz  $\mathbf{P}$  contiene las cargas, o los espectros ocultos/estructuras comunes, la matriz  $\mathbf{T}$  la concentración de cada una de las cargas correspondiente a cada muestra y la matriz  $\mathbf{E}$  contiene los valores residuales (la parte que el modelo no describe (consulte [2] para una descripción más detallada del ACP aplicado al espectro NIR)).

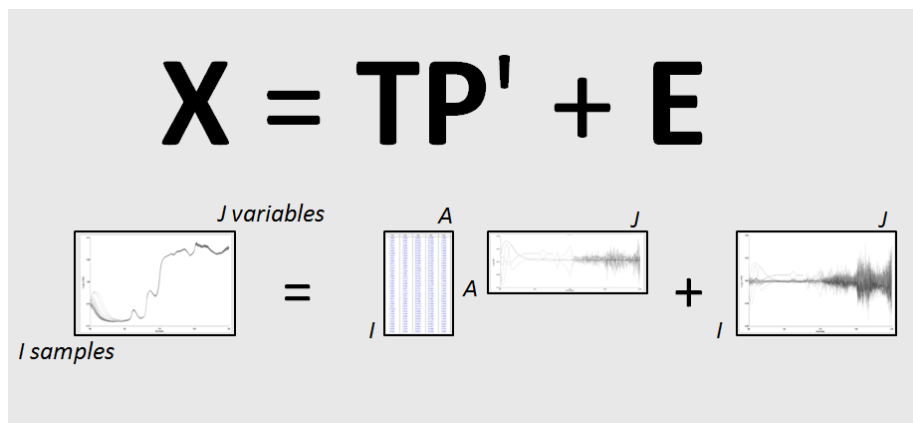


Figura 1. El modelo ACP.

Existen dos formas de analizar la distancia de una muestra concreta a otra: puede medirse como la distancia hasta el centro *dentro* del modelo (se calcula a partir de los resultados), o se pueden calcular las distancias *hasta* el modelo (calculadas a partir de los valores residuales):

- El espectro de una muestra correspondiente a una muestra nueva que presenta una desviación sólo respecto al espacio de los resultados puede denominarse como un *valor atípico de concentración* (el patrón del espectro es comparable a los espectros de un buen producto, pero la muestra o bien contiene concentraciones muy altas o muy bajas, en comparación con las muestras de buenos productos).
- El espectro de una muestra correspondiente a una muestra nueva que presenta una desviación sólo respecto a los valores residuales puede denominarse como un *valor atípico del patrón del espectro* (el patrón espectral presenta una desviación en comparación con los espectros de un buen producto, pero los niveles de concentración de la muestra son comparables a las muestras de buenos productos).
- Un espectro que presenta una desviación respecto a ambas distancias (resultados y valores residuales) es una concentración y un valor atípico del patrón espectral.

Los métodos de cualificación WinISI están diseñados para caracterizar a un conjunto de espectros que representen muestras de productos que se ajusten a las especificaciones de calidad y para detectar desviaciones del conjunto de datos. Están disponibles las siguientes distancias:

- **Distancia de Mahalanobis (al cuadrado):** mide la distancia entre un espectro y el espectro medio del conjunto de datos del buen producto mediante resultados del ACP ( $\mathbf{T}$ ). Los valores se colocan en una escala, de tal forma que la distancia media para todas las muestras del ACP sea de 1,0. Esta medición resulta útil para detectar si el espectro de una muestra nueva cercano al espectro medio se asemeja del algún modo a las muestras de los productos buenos.

- **Distancia vecina de Mahalanobis (al cuadrado):** mide la distancia entre un espectro y el espectro *más cercano* en el conjunto de datos de los productos buenos. Los valores se colocan en una escala utilizando el mismo factor utilizado en la medición de la distancia de Mahalanobis. Esta medición resulta útil para detectar la distancia al espectro vecino más cercano y compararla con la distancia vecina media respecto a la totalidad del conjunto de datos.
- **Valores residuales RMS X:** la parte de un espectro que no se puede explicar con el modelo ACP. Se calcula la raíz cuadrada de la media aritmética de los cuadrados del valor residual (RMS, o media cuadrática) para obtener los valores residuales RMS X. Esta medición resulta útil cuando se quiere detectar una variación espectral que no sea similar a la variación espectral en el conjunto de datos de los productos buenos.
- **Valor residual Máximo X:** En lugar de calcular el RMS, este método obtiene el valor residual absoluto más grande X correspondiente al valor residual del espectro de la nueva muestra. Es más sensible a la variación sin modelar con tan sólo unas longitudes de onda que el método de valores residuales RMS X. Si, por ejemplo, un adulterante tiene un pico específico a 2.100 nm y a 2.300 nm, podría resultar más sencillo detectarlo centrándose en la parte específica del espectro en la que la desviación es mayor.
- **Valor residual T Máximo X:** es una modificación del método del Valor residual Máximo X. Cada longitud de onda residual X se convierte en una estadística "t" dividiéndola entre la desviación estándar de todos los valores residuales correspondientes para esa longitud de onda en el conjunto de productos buenos. El método calcula el valor "t" absoluto más alto. Este método es más sensible a pequeñas desviaciones de los valores residuales, en comparación con los valores residuales en el conjunto de datos de productos buenos.
- **Pico Máximo T:** este método también está basado en los valores residuales X de un modelo de ACP, al igual que los tres métodos anteriores. El método busca una forma de pico en los valores residuales X. Este método calcula una estadística "t" dividiendo la intensidad de cada pico entre la desviación estándar de las intensidades de todos los picos correspondientes. El método calcula el valor positivo "t" más alto en todas las longitudes de onda. Este método es sensible a la adición de nuevos ingredientes que no estén presentes en el conjunto de datos de productos buenos y reaccionará ante desviaciones espectrales.
- **Método ASM:** este método es una combinación de la distancia de Mahalanobis y de los valores residuales RMS X. Es la raíz cuadrada de la suma de la distancia de Mahalanobis al cuadrado y unos valores residuales RMS X debidamente colocados en una escala al cuadrado. Esta medición indicará que una muestra es buena sólo si el espectro es similar al espectro medio de un producto bueno y los valores residuales X son reducidos. De esta forma, los métodos ASM combinan información procedente tanto del modelo como de fuera del modelo.

Los métodos basados en Mahalanobis resultan adecuados para la detección de valores atípicos, mientras que los métodos basados en los valores residuales resultan pertinentes para la detección de una variedad de patrones espectrales; el método ASM fusiona la concentración y la detección de valores atípicos en los patrones espectrales.

## Estudio: los datos

Se registraron los espectros NIR de 214 muestras de leche desnatada en polvo mediante un instrumento NIRS DS2500 con el fin de proporcionar espectros de 400 a 2.500 nm con intervalos de 0,5 nm (en total, 4.200 variables por espectro (Figura 2)). Se dividieron en un conjunto de instrucción de 81 muestras, un conjunto de ajuste (o prueba) de 40 muestras y un conjunto de validación de 93 muestras. Los subgrupos se hicieron a partir de las 214 muestras ordenadas por fecha de análisis, siendo el conjunto de instrucción el más antiguo, el conjunto de ajuste ocupando una posición media y el conjunto de validación correspondiendo a los resultados más recientes. Asimismo, se mezclaron los espectros de ajuste y validación con el 0,5% de melamina (Figura 3). Dado que ni la leche desnatada en polvo ni la melamina presentan picos distintivos en el rango de 400-1.100 nm, en el estudio sólo se utilizó el rango NIR.

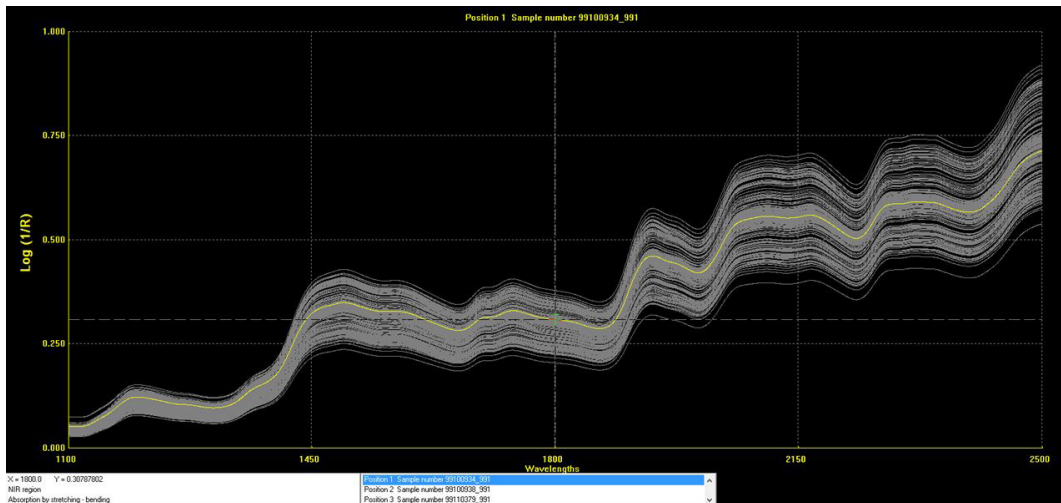


Figura 2. Espectros NIR del rango 1.100-2.500 nm de las 214 muestras de leche desnatada en polvo.

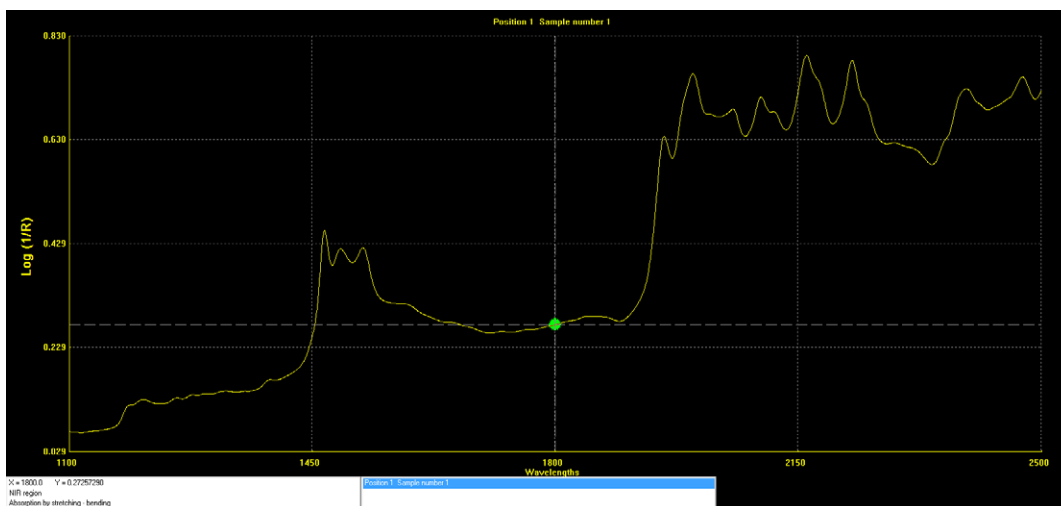


Figura 3. Espectro NIR de la melamina del rango 1.100-2500 nm.

Las mismas condiciones utilizadas para la selección de muestras en el análisis cuantitativo resultan válidas aquí (es importante incluir la variación esperada que aparecerá en futuras muestras "buenas"). Se puede optar por una variación limitada (por ejemplo, en el contexto de un proceso, el último mes de producción o, para soluciones de laboratorio, una variación más amplia).

## El modelo de Calificación

Nos interesa detectar la melamina que presenta picos claros alrededor de 1.467 nm y varios entre 1.955 y 2.227 nm; además, se preprocesó con un segundo derivado (2, 20, 1, 1) para resaltar estos picos.

Se desarrolló un modelo de calificación (Definición de producto bueno -GPD, por sus siglas en inglés- en WinISI) a partir de las 81 muestras de instrucción. El resultado se muestra en la Figura 4. La parte difícil consiste en determinar el número de componentes que hay que utilizar en el modelo. Para ello, utilizamos el conjunto de ajuste de 40 muestras y algunas muestras mezcladas.

Tras comparar los siete métodos disponibles para llevar a cabo la calificación, resultó obvio que el método de Pico Máximo T era muy eficaz a la hora de detectar la melamina, por lo que basaremos el análisis en este método.

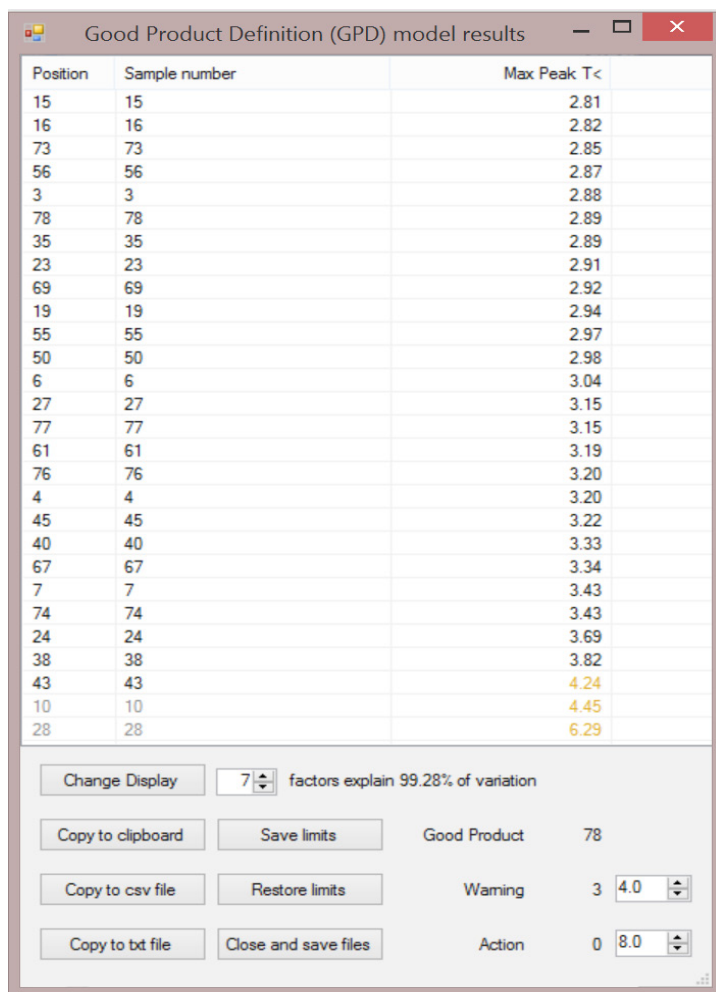


Figura 4. Resultado de la aplicación del modelo de calificación.

Si el número de factores del modelo de calificación es demasiado bajo, la creación de modelos para el producto será incompleta (infra-ajuste) y los constituyentes de las muestras naturales pueden contribuir a obtener un valor alto de Pico Máximo T. Si el número de factores es demasiado alto (sobre-ajuste), el ruido aleatorio puede contribuir a obtener un valor alto del Pico Máximo T. En la Tabla 1 se recogen los valores mínimos y máximos del Pico Máximo T correspondientes al conjunto de ajuste y al conjunto de ajuste mezclado. Es aconsejable que el valor mezclado mínimo sea mayor que el valor máximo sin mezclar. Si se utiliza el ratio de valor mínimo mezclado dividido entre el valor máximo sin mezclar, los siete factores hacen un trabajo excepcional a la hora de identificar la melamina.

Factores	Conjunto de ajuste		Conjunto de ajuste mezclado	
	Mín.	Máx.	Mín.	Máx.
3	1,97	7,44	12,57	17,43
4	2,09	7,87	12,40	17,92
5	2,36	7,87	12,28	20,76
6	3,08	7,69	51,13	58,27
7	3,09	7,43	51,69	58,75
8	3,45	7,88	48,56	56,91
9	3,59	7,90	48,91	56,93
10	3,27	7,58	47,26	55,54

Tabla 1. Rango de los valores de Pico Máximo T correspondientes al conjunto de ajuste y al conjunto de ajuste mezclado.

Debería elegirse un valor de umbral ligeramente superior al máximo sin mezclar observado. En este caso 8,0 sería un umbral adecuado (para este valor, el número de falsos positivos debería registrarse también para un número lo suficientemente grande de muestras en blanco). Cuando se aplica el modelo de calificación de siete factores al conjunto de validación, obtenemos un rango sin mezclar de 2,27-7,79 y un rango mezclado de 50,74-59,44, identificando claramente la melamina en los espectros.

Si hubiese que configurar un análisis general de la adulteración con NIR, en el software Mosaic-Nova se pueden combinar, por ejemplo, a) la distancia de Mahalanobis para detectar los valores atípicos en las concentraciones, b) valores residuales RMS X para detectar valores atípicos de patrones espectrales amplios y c) Pico Máximo T para detectar valores atípicos de patrones espectrales estrechos y específicos.

## Conclusión

Gracias a las funciones del nuevo WinISI, resulta sencillo desarrollar modelos de cualificación para la plataforma NIRS DS2500; los modelos se integran a la perfección en un flujo de IQ<sup>2</sup> con *Identificación* (selección del grupo de productos adecuado, por ejemplo, leche desnatada en polvo), *Calificación* (estamos ante un producto "bueno" que no está adulterado) y *Cuantificación* (predicción de proteínas y humedad) y ofrecen una herramienta muy sólida para garantizar la obtención de un buen producto.

## Referencias

[1] Chemometric terminology for qualitative and quantitative analysis – IQ<sup>2</sup>, InFocus Vol. 38, N.º 1, 2014.

[2] Principal Component Analysis and Near Infrared Spectroscopy, In Focus Vol. 36, N.º 1, 2012 (<http://viewer.zmags.com/publication/b452b1b8#/b452b1b8/33>)

# FOSS

FOSS  
Avda. Josep Tarradellas, 8-10  
08029 - BARCELONA  
España

Tel.: 934 949 940  
Fax: 934 052 176

infofoss@foss.es  
[www.foss.es](http://www.foss.es)