

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 537 000**

51 Int. Cl.:

G01N 21/64 (2006.01)

G01N 33/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **08.12.2010 E 10805253 (1)**

97 Fecha y número de publicación de la concesión europea: **18.02.2015 EP 2513632**

54 Título: **Procedimiento y aparato de análisis espectroscópico, en particular de alimentos, que emplea un tratamiento multivía de los datos espectroscópicos**

30 Prioridad:

26.03.2010 FR 1001221

16.12.2009 FR 0906088

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

01.06.2015

73 Titular/es:

**SPECTRALYS INNOVATION (100.0%)
Biocitech 102 avenue Gaston Roussel
93230 Romainville, FR**

72 Inventor/es:

**RIZKALLAH, JAD y
BIRLOUEZ-ARAGON, INÈS**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 537 000 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimiento y aparato de análisis espectroscópico, en particular de alimentos, que emplea un tratamiento multivía de los datos espectroscópicos

5 La invención trata sobre un procedimiento y un aparato de análisis espectroscópico de al menos una muestra, en particular de un alimento o medicamento, que emplea un método de análisis de datos espectroscópicos basado en un modelo estadístico multivía. En particular, la invención se puede aplicar al estudio de la evolución de las propiedades nutricionales (contenido en vitaminas...) y/o toxicológicas (contenido en contaminantes de neoformación...) de un alimento en el curso de su preparación o conservación, pero no solamente a ello. De manera más general, la invención se puede aplicar a la determinación de cualquier indicador de calidad de una muestra y/o

10 de cualquier parámetro que caracteriza un proceso al cual se somete dicha muestra.

Los contaminantes de neoformación (CNF) son moléculas nuevas formadas en un alimento a partir de componentes propios de ese alimento, debido a la acción de procesos de transformación y que presentan problemas de salud pública. Pocos laboratorios públicos o privados son capaces hoy día de determinar la dosis de estos contaminantes. Además, si bien se han publicado varios métodos para medir la cantidad de cada uno de estos contaminantes,

15 ninguno de ellos ha sido objeto, realmente, de un proceso de estandarización y normalización. Por otra parte, estos análisis son costosos y el plazo de obtención de los resultados es largo (2 a 3 semanas). La industria necesita métodos de análisis más rápidos y menos costosos para asegurar un control de la conformidad de sus procedimientos y de la seguridad de los alimentos que produce.

Los procesos de transformación alimentaria (cocción, esterilización, conservación) pueden también influir de manera desfavorable en las propiedades nutricionales de los alimentos, por ejemplo disminuyendo su cantidad de vitaminas. La caracterización de estos procesos y de su influencia sobre las propiedades nutricionales de los alimentos presenta problemas análogos a los citados previamente de determinación de cantidades de los contaminantes de neoformación.

20

Se sabe realizar análisis, en especial espectroscópicos, de los alimentos recurriendo a métodos de la quimiometría y, en particular, al análisis multivía. El análisis multivía es la ampliación natural del análisis multivariante cuando se disponen los datos en tablas de tres o más vías. Se basa en la utilización de modelos estadísticos tales como "PARAFAC" ("Parallel Factor", es decir, modelo de factores paralelos) y NPLS ("Regresión de mínimos cuadrados parciales de n-vías", por sus siglas en inglés). Estos métodos, así como su aplicación al análisis de productos alimentarios, se describen en el documento [Bro 1998].

25

Más precisamente, [Rizkallah 2007] describe un método de análisis de muestras no preparadas (solamente molidas, si es necesario) basado en la aplicación del modelo PARAFAC a espectros de fluorescencia frontal ("front-face fluorescence"). Este método conlleva la iluminación de una muestra por varios rayos luminosos monocromáticos de longitudes de onda respectivas y la adquisición de los espectros de fluorescencia correspondientes. Las radiaciones de excitación muestrean finamente (varios centenares de puntos) una zona espectral que cubre el visible y el UV próximo; a su vez, los espectros de emisión de fluorescencia son muestreados espectralmente (asimismo sobre varios centenares de puntos). Los datos espectrales así recogidos para cada muestra se disponen en una matriz de grandes dimensiones denominada "matriz de excitación y emisión" (MEE), en la cual una dimensión representa las longitudes de onda de excitación y la otra dimensión representa las longitudes de onda de emisión.

30

35

Las MEE que corresponden a varias muestras de calibrado son analizadas mediante el método PARAFAC el cual permite extraer informaciones, denominadas "factores PARAFAC" que corresponden a los perfiles bilineales de fluorescencia y a sus intensidades relativas. Luego, una regresión múltiple entre las intensidades de fluorescencia y las concentraciones de contaminantes de neoformación (o de otras sustancias de interés) medidas químicamente permiten construir un modelo de calibrado que se utiliza a continuación para predecir las concentraciones de contaminantes de neoformación a partir de las MEE recogidas de nuevas muestras para analizar.

40

45 Este método presenta dos limitaciones mayores:

- En primer lugar, necesita un equipamiento de elevado coste: un espectrofluorímetro equipado con una lámpara de xenón que proporciona el conjunto de las longitudes de onda desde el UV próximo hasta el visible y dos monocromadores capaces de hacer pasar estas diferentes longitudes de onda tanto en excitación como en emisión para los fotones emitidos por la muestra.
- 50 - En segundo lugar, su puesta en marcha es larga ya que, incluso si se evita la preparación de la muestra, la adquisición de cada EEM necesita de 15 a 45 minutos, según la resolución espectral buscada. El análisis de estas matrices de datos, que comprenden un número considerable de variables, es igualmente muy pesado, a pesar de la automatización de los algoritmos de tratamiento de datos.

Métodos similares son descritos por [Rizkallah et al., 2008] y [Christensen et al. 2005].

55 El documento [Nahorniak 2003] estudia la influencia de la resolución espectral (y por tanto del tamaño de la matriz EEM) sobre el error de predicción de la concentración de un fluoróforo por aplicación del método PARAFAC. En ese

documento se consideran disoluciones diluidas y de composición simple y conocida; además, todas las medidas se efectúan con una resolución espectral elevada, de 2 nm/píxel, siendo obtenida la reducción de resolución calculando medias espectrales.

5 Resulta, de este estudio, que se minimiza el error de predicción –y vale aproximadamente 0,5–1 %- para una resolución en excitación de 6 a 10 nm/píxel; luego aumenta fuertemente para resoluciones inferiores (aumento de un factor 2 a 6 pasando a una resolución de 20 nm/píxel). Parece por tanto difícil utilizar resoluciones inferiores a 20 nm/píxel.

La invención aspira a proporcionar un método de análisis más simple, más rápido, que necesite un equipo más sencillo y menos costoso y, en consecuencia, mejor adaptada a las exigencias industriales.

10 De acuerdo con la invención, se alcanza tal objetivo mediante un procedimiento de análisis espectrocópico según la reivindicación 1.

De manera ventajosa, el número de radiaciones luminosas de excitación y de espectros de fluorescencia correspondientes para cada muestra, puede estar comprendido entre dos y seis y, preferentemente, entre tres y cinco, en una zona espectral de al menos 100 nm y, preferentemente, de al menos 150 nm.

15 Este método es innovador debido a que aplica herramientas de análisis multivía como el modelo PARAFAC, normalmente utilizado para MEE de grandes dimensiones, a un conjunto constituido por un número limitado de espectros (generalmente dos a seis, preferentemente de tres a cinco), muestreando de forma muy tosca la zona espectral de excitación de las muestras. Esta zona es relativamente amplia (al menos 100, incluso 150 nm, o más) porque la composición físicoquímica de dichas muestras es ampliamente desconocida y por tanto la posición exacta de los espectros de absorción de los fluoróforos contenidos en dichas muestras también lo es.

20 Los presentes inventores han descubierto, de manera sorprendente, que los modelos multivía continúan siendo aplicables y proporcionan resultados explotables incluso utilizando una resolución espectral en excitación tan tosca y un número tan restringido de longitudes de onda de excitación.

25 Es importante subrayar que, desde un punto de vista físico, los factores de carga o “loadings” representan los espectros de emisión o de excitación de cada factor y las puntuaciones factoriales o “scores” las intensidades de dichos factores. Un “factor PARAFAC” representa la contribución de un fluoróforo, o de una mezcla de fluoróforos, común a las muestras analizadas. Es preciso subrayar también que los espectros de fluorescencia son deformados por la interacción entre los fotones de excitación y los fotones emitidos con la matriz alimentaria y que no se satisface la condición de trilinealidad generalmente admitida para la aplicación del modelo PARAFAC. No obstante, los inventores han mostrado que este modelo se puede aplicar de forma útil a tales matrices EEM para describir la composición en fluoróforos.

30 Es necesario comprender bien que el procedimiento de la invención no pretende (al menos, no necesariamente) determinar el contenido de los propios fluoróforos. Este procedimiento explota simplemente correlaciones, en general no lineales o multilineales, entre las evoluciones de los espectros de excitación y de emisión de fluorescencia y ciertos parámetros de calidad de la muestra (contenido de nutrientes y/o de contaminantes que, la mayoría de las veces, no son fluorescentes; carga microbiana,...) o ciertos parámetros que caracterizan un procedimiento al cual dicha muestra ha sido sometida (valor de esterilización, de pasteurización o de cocción, o un producto tiempo-temperatura, por ejemplo).

35 Es especialmente interesante una comparación con el documento citado previamente [Nahorniak 2003]. Este documento se refiere a un caso especialmente favorable, a saber, el análisis de disoluciones de composición simple y conocida. Por el contrario, la invención se refiere al análisis por fluorescencia frontal de muestras complejas y fuertemente difusoras, cuya composición prácticamente no se conoce. Además, en el documento [Nahorniak 2003] los espectros se toman con alta resolución espectral de excitación y luego se promedian, lo que por sí tiende a reducir el error de predicción. Por el contrario, en el caso de la invención, solamente se toman de dos a seis espectros, con una resolución espectral en excitación muy baja, inferior a 20 nm/píxel e incluso a 50 nm/píxel. A pesar de esto, una elección oportuna de los parámetros del modelo multivía y/o un tratamiento previo oportuno de los datos permiten limitar el error de predicción a valores completamente aceptables, típicamente inferiores a 15 % e incluso 10 %.

Un procedimiento según la invención puede tener asimismo una fase preliminar de calibrado que consta de:

- 50
- i) la iluminación de varias (dos o más) muestras de calibrado por dichas radiaciones luminosas de excitación;
 - ii) la adquisición de los espectros de fluorescencia frontal de las muestras, que corresponden a dichas radiaciones luminosas de excitación;
 - iii) un tratamiento previo de los espectros de fluorescencia tomados;

- iv) la determinación, mediante un método iterativo, de dichos vectores de los factores de carga (“loadings”) de excitación y de emisión del modelo multivía, así como de un vector de puntuaciones factoriales (“scores”) para cada muestra de calibrado; y
- v) la determinación de una función de regresión que liga dichos vectores de puntuaciones a los valores, conocidos de dicho o de cada parámetro, para dichas muestras de calibrado.

Los espectros de fluorescencia pueden ser espectros de autofluorescencia o, en ciertos casos, espectros de fluorescencia inducida por un marcador añadido a la muestra. Por ejemplo, la estimación de la carga microbiana de una muestra se puede facilitar mucho utilizando una sonda fluorescente.

De manera ventajosa, los espectros de fluorescencia pueden ser tratados previamente mediante sustracción de una contribución debida a la difusión Rayleigh de primer orden de la radiación luminosa de excitación, siendo calculada dicha contribución por medio de un modelo lineal generalizado. En particular, dicho modelo lineal generalizado se puede determinar a partir de una primera parte o zona del espectro a tratar previamente, que comprende solamente dicha contribución debida a la difusión Rayleigh y es utilizada para predecir la contribución de la difusión Rayleigh en una segunda parte o zona de dicho espectro en la cual se superpone a una contribución de fluorescencia. La sustracción de la difusión Rayleigh es necesaria también en los procedimientos de análisis conocidos de la técnica anterior, pero se hace crítica en el procedimiento de la invención. La utilización de un modelo lineal generalizado, en lugar de técnicas conocidas de sustitución por ceros o valores perdidos, resulta ser especialmente ventajosa cuando se utiliza un número muy bajo de longitudes de onda de excitación, como en el caso de la invención.

A continuación, los espectros de fluorescencia se pueden tratar previamente normalizándolos. Varias muestras, de calibrado y/o para analizar, se pueden también tratar previamente efectuando una corrección multiplicativa de dispersión (MSC, por las siglas en inglés de “multiplicative scatter correction”).

Según un modo de realización preferido de la invención, dicho modelo estadístico multivía puede ser un modelo PARAFAC, estando representados los espectros de fluorescencia de cada muestra en forma concatenada. En ese caso, los vectores de los factores de carga de dicho modelo PARAFAC se pueden determinar mediante un método iterativo que se detiene cuando la disminución de la función de pérdida debida a la última iteración desciende por debajo de un valor umbral (“parámetro de convergencia”) comprendido entre 10^{-4} y 10^{-2} . Este valor debe ser comparado al utilizado generalmente de 10^{-6} . En efecto, la utilización de un parámetro de convergencia más grande facilita la convergencia del modelo PARAFAC cuando se utiliza un número bajo de longitudes de onda de excitación.

Como variante, dicho modelo estadístico multivía puede ser un modelo NPLS, estando representados los espectros de fluorescencia de cada muestra de forma concatenada. En este caso, los espectros de fluorescencia pueden ser tratados previamente de forma útil por corrección ortogonal de la señal.

De forma ventajosa, dichas radiaciones luminosas de excitación pueden comprender: una primera radiación de una longitud de onda comprendida entre 270 y 300 nm; una segunda radiación de una longitud de onda comprendida entre 300 y 360 nm; y una tercera radiación de una longitud de onda comprendida entre 400 y 500 nm.

Dicho parámetro o al menos uno de los que se van a determinar se puede escoger, en especial, entre: el contenido de una sustancia de interés en dicha muestra (o en cada una de ellas); una carga microbiana de dicha muestra o de cada una de ellas; y un valor de esterilización, de pasteurización o de cocción, o un producto tiempo-temperatura, de un procedimiento al cual dicha muestra (o cada una de ellas) ha sido sometida.

Las nociones de valor de esterilización, de pasteurización o de cocción son bien conocidas de suyo. Se considera una etapa de un proceso térmico en el curso de la cual la muestra se mantiene a una temperatura T_i (en °C, en el núcleo de la muestra) durante un tiempo t_i . El valor de pasteurización de esta etapa viene dado por:

$$VP_i = t_i \times 10^{(T_i - 70)/Z};$$

su valor de esterilización viene dado por:

$$VS_i = t_i \times 10^{(T_i - 121,1)/Z};$$

y su valor de cocción está dado por:

$$VC_i = t_i \times 10^{(T_i - 100)/Z};$$

donde el parámetro Z depende del microorganismo considerado (para los valores de pasteurización y de esterilización) o de la naturaleza de la muestra (para el valor de cocción).

Para un proceso multi-etapas se toma la suma de los valores de pasteurización o de esterilización de las diferentes etapas.

Las muestras para analizar y de calibrado pueden ser, especialmente, muestras de un producto escogido entre un alimento y un medicamento.

Otro objeto de la invención es la utilización de un procedimiento tal como se describe previamente para medir la evolución de las propiedades nutricionales, microbiológicas y/o toxicológicas de un alimento durante el curso de su preparación o conservación.

5 Otro objeto más de la invención es un aparato de análisis espectroscópico según la reivindicación 13. Preferentemente, tal aparato puede tener entre dos y seis (y preferentemente entre tres y cinco) de dichas fuentes luminosas.

10 De manera ventajosa, un aparato tal puede tener: una primera fuente luminosa que emite una radiación de una longitud de onda comprendida entre 270 y 300 nm; una segunda fuente luminosa que emite una radiación de una longitud de onda comprendida entre 300 y 360 nm; y una tercera fuente luminosa que emite una radiación de una longitud de onda comprendida entre 400 y 500 nm.

Otras características, detalles y ventajas de la invención resaltarán en la lectura de la descripción de este documento realizada tomando como referencia los dibujos anexos, dados a título de ejemplo, y que representan, respectivamente:

- 15 - la figura 1, tres espectros “en bruto” de fluorescencia de una muestra de achicoria torrefactada, que corresponde a tres radiaciones de excitación a tres longitudes de onda diferentes;
- las figuras 2A, 2B y 2C, tres gráficas que ilustran la operación de sustracción de la difusión de Rayleigh por el método que utiliza un modelo lineal generalizado;
- 20 - las figuras 3A, 3B y 3C, tres gráficas que comparan dicho método de sustracción de la difusión de Rayleigh (figura 3C) con otros dos métodos conocidos de la técnica anterior: la sustitución por ceros (figura 3A) y la sustitución por valores perdidos (figura 3B);
- la figura 4, una ilustración esquemática de la noción de concatenación de los espectros;
- la figura 5, una ilustración esquemática del método PARAFAC;
- 25 - la figura 6, el esquema de principio de un aparato de análisis según un modelo de realización de la invención;
- las figuras 7 y 8, gráficos que ilustran la aplicación de un método según un primer modo de realización de la invención, basado en un modelo PARAFAC, para la predicción de la cantidad de acrilamida en diferentes muestras de achicoria;
- 30 - las figuras 9A, 9B, 10A y 10B, gráficas que ilustran la aplicación de un método según un segundo modo de realización de la invención, basado en un modelo NPLS, para la predicción del contenido de vitamina C de diferentes muestras de judías verdes.

35 El procedimiento de la invención utiliza la señal de fluorescencia emitida en la superficie del alimento después de iluminación por haces luminosos de longitudes de onda determinadas del campo UV – visible (aproximadamente: 250 – 750 nm). Esta señal es analizada por métodos quimiométricos que permiten extraer la información que se encuentra correlacionada con los parámetros que se busca medir. La existencia de tal correlación se deduce del hecho de que, en el curso de la producción agrícola, de la conservación y de la transformación, la fluorescencia intrínseca de los constituyentes naturales del alimento (vitaminas, proteínas y otros constituyentes naturales o añadidos de manera intencionada o no intencionada), así como su reflectancia, evolucionan, mientras que, al mismo tiempo, aparecen nuevas señales debido a la formación de nuevas moléculas (entre ellos los contaminantes de neoformación). Se habla de fluorescencia de neoformación o adquirida según que los fluoróforos se formen *ex novo* o provengan del medio ambiente. La evolución conjunta de las señales nativas (SN), de neoformación (SNF) y adquiridas de nuevas (SNA) se correlaciona de forma fuerte con las modificaciones físicas, físico-químicas o microbiológicas del alimento, en particular con los cambios de los parámetros de calidad, inducidos durante el curso de la producción, la conservación y/o la transformación. Los factores de cambio que influyen sobre la calidad del alimento son las radiaciones ultravioletas oxidantes, la destrucción de microorganismos o, al contrario, el desarrollo de algunos de ellos que pueden iniciar la síntesis de micotoxinas, la intervención humana sobre los cultivos (abonos, pesticidas...) o la aplicación de procesos que modifican la temperatura, la presión o cualquier otro parámetro físico en el seno del alimento y que inducen en consecuencia una modificación de la composición físico-química y de los parámetros de calidad.

50 Las radiaciones luminosas de excitación presentan longitudes de onda escogidas de forma que se explore el espectro UV – visible de la forma más amplia posible. De manera general, se pueden escoger a priori:

- una longitud de onda comprendida entre 270 y 300 nm que permite excitar el triptófano, los fenoles como el ácido clorogénico o el hidroxitirosol o incluso la vitamina E, moléculas que emiten en el UV;

- una longitud de onda comprendida entre 400 y 450 o 500 nm que permite excitar la riboflavina, las porfirinas y la clorofila, moléculas que emiten en el visible (500 – 700 nm);
- una o dos longitudes de onda comprendidas entre 300 y 360 e incluso 400 nm se introducen para excitar los fluoróforos de neoformación, esencialmente productos de Maillard y productos de peroxidación lipídica, pero también las micotoxinas que emiten en el UV lejano-próximo visible (400 – 500 nm)

Estas longitudes de onda serán modificadas en caso de aplicaciones distintas de las de la industria alimentaria. En todo caso, para precisar más, es posible escoger las longitudes de onda más cercanas posibles a los máximos de los factores de carga que representan el vector de excitación obtenido por la descomposición PARAFAC de una matriz MEE completa obtenida con un fluorímetro de laboratorio sobre un lote de muestras representativas.

El número de longitudes de onda utilizadas está comprendido entre 2 y 6, de forma ventajosa entre 3 y 5 y preferentemente es igual a 5, lo que permite excitar cada vez un número correspondiente de grupos de fluoróforos entre los descritos antes en el texto. La intensidad de la radiación de excitación se escoge de suerte que la energía de emisión de fluorescencia de estos fluoróforos sea modificada de manera significativa en el curso de las etapas de transformación que se busca caracterizar. Por ejemplo, la intensidad debe ser suficiente para que la aparición de micotoxinas ligada al desarrollo de hongos en un grano en el curso de su transformación modifique sensiblemente la emisión de fluorescencia recogida en el detector a partir de una muestra no contaminada. Debe asimismo poner en evidencia la aplicación de tal tratamiento térmico caracterizado por una pareja tiempo – temperatura; o incluso la aplicación de un lavado, blanqueamiento o tratamiento de descontaminación.

La figura 6 muestra un esquema muy simplificado de un aparato para poner en marcha el procedimiento de la invención. Este aparato comprende tres fuentes luminosas S_1 , S_2 y S_3 que emiten cada una un haz de radiación monocromática a una longitud de onda diferente dirigida de forma que se ilumine la muestra E. La señal de fluorescencia F (de hecho, mezcla de fluorescencia y difusión Rayleigh de primer y, a veces, de segundo orden) emitida por esta muestra es transportada por una fibra óptica hacia un espectrómetro que descompone la radiación luminosa emitida en un espectro. Los espectros tomados se tratan mediante un medio de tratamiento MTD (típicamente, un ordenador programado de una forma oportuna) que permite extraer la información quimiométrica buscada.

Típicamente, las fuentes luminosas pueden ser diodos electroluminiscentes, incluso láseres (preferentemente de semiconductor) si se necesitan intensidades mayores.

El análisis de los datos espectroscópicos, efectuado por el ordenador MTD comprende cuatro etapas principales:

- el tratamiento previo de los espectros;
- la determinación de los vectores de los factores de carga (“loadings”) del modelo multivía con la ayuda de los espectros de las muestras de calibrado;
- la determinación de los vectores de las puntuaciones factoriales (“scores”) aplicando ese modelo a los espectros de las muestras que se van a analizar; y
- la determinación de una información de naturaleza química (contenido de una o de varias sustancias de interés), física (producto tiempo – temperatura, valor de esterilización, de pasteurización o de cocción), físico-química o microbiológica (carga microbiana) a partir de dichos vectores de las puntuaciones factoriales.

Se considera en primer lugar el tratamiento previo de los espectros, tomando como referencia un ejemplo específico, en el cual una muestra de achicoria torrefactada se ilumina de manera sucesiva por tres radiaciones de excitación a 280, 340 y 429 nm. Cada uno de los tres espectros de fluorescencia está constituido por 1515 valores de intensidad espectral para otras tantas longitudes de onda λ diferentes. la resolución espectral es de 0,25 nm/pixel, pero puede dividirse por 5, incluso por más, sin que se observe una degradación de los resultados del procedimiento de la invención.

Los espectros “en bruto” (figura 1) están dominados por la difusión de Rayleigh de primer orden, a la longitud de onda de las radiaciones de excitación (280, 340 y 429 nm). El espectro de cada señal de fluorescencia, en efecto, está parcialmente superpuesto a la radiación de excitación difundida, que es mucho más intensa. Este problema es conocido y generalmente se resuelve reemplazando los valores de intensidad que corresponden a la región espectral de superposición por ceros (figura 3A) o por valores perdidos (figura 3B). En estas figuras, las referencias F_{280} , F_{340} y F_{429} representan los espectros de fluorescencia relativos a las excitaciones a 280, 340 y 429 nm, respectivamente, en las cuales las intensidades de las regiones espectrales de superposición con la difusión de Rayleigh se han reemplazado por ceros o por valores perdidos.

Estos métodos presentan inconvenientes, ya que la sustitución por ceros engendra “artefactos” y, en consecuencia, falsea el análisis creando varianzas artificiales. La sustitución por valores perdidos produce dificultades en cuanto a la convergencia de los modelos multivía utilizados para el análisis de datos. Además, la presencia de los valores

perdidos impide la utilización de tratamientos previos de estandarización (véase más adelante) que no se pueden aplicar más que a valores reales.

5 Estos problemas son aceptables en el caso de las técnicas convencionales, basadas en el empleo de MEE de grandes dimensiones, pero resultan completamente inaceptables en un procedimiento tal como el de la invención, basado en el análisis de datos muy poco resueltos en longitud de onda de excitación.

10 Por esta razón, la invención propone suprimir la contribución de la difusión de Rayleigh de primer orden mediante una técnica innovadora que se basa en la predicción de la región de difusión que se superpone con la fluorescencia a través de un modelo lineal generalizado (GLZ) con una función de enlace log [Davidson 1998; Rizkallah 2007]. Este modelo se calibra en una región del espectro en la cual la contribución de la fluorescencia es desdoblable y la intensidad espectral se atribuye de manera exclusiva a la difusión (referencia RD en las figuras 2A y 2B).

El modelo lineal generalizado es de la forma:

$$f(\mu_y) = b_0 + b_1x$$

15 En la ecuación, $f(\mu_y)$ es la función de enlace de μ_y , valor esperado de y , siendo y el vector de las intensidades de difusión de Rayleigh que no se superponen con la fluorescencia, mientras que x es un vector de índices (1, 2, 3, etc) del mismo tamaño que y .

20 Gracias al modelo generalizado, se pueden modelizar mediante la función de enlace relaciones no lineales (entre x e y). El modelo generalizado se puede utilizar para modelizar variables dependientes que tienen distribuciones que pertenecen a la familia exponencial (distribuciones Normal, Gamma, Poisson, etc). La regresión lineal múltiple es un caso especial del modelo GLZ que corresponde a una función de enlace igual a la función identidad y a una variable dependiente (y) que presenta una distribución normal.

Los parámetros b_i del modelo GLZ se estiman mediante el método estadístico de maximización de la probabilidad (L):

$$L = F(Y, modelo) = \prod_{i=1}^n p[y_i, b_i]$$

siendo $p[y_i, b_i]$ la probabilidad de y_i condicionada por b_i .

25 El objetivo es encontrar los parámetros que dan la mayor probabilidad (densidad conjunta) de realización de y para todas las observaciones. Se utiliza una estimación iterativa (algoritmo de Fisher, que es un método casi-newtoniano) para encontrar los parámetros b_i maximizando L:

$$\frac{\partial \log(L)}{\partial b_i} = 0$$

30 Una vez estimados los b_i , se aplica el modelo GLZ sobre los índices de la difusión que corresponden a la región espectral superpuesta con la fluorescencia (figura 2B, región de predicción RP), para predecir las intensidades de la difusión "pura". Tras esta predicción, se sustraen los espectros de difusión completos (parte real y predicha) de los MEE para obtener el espectro de fluorescencia puro SF (figura 2C, a comparar con la figura 2A que muestra el espectro de fluorescencia parcialmente superpuesto al espectro de la luz difundida).

35 Las sucesivas operaciones no se realizan solo sobre los tres espectros de fluorescencia considerados de forma individual, sino sobre los espectros concatenados, es decir, dispuestos uno a continuación de otro en una columna única (véase la figura 4).

Según el caso, los espectros concatenados se pueden someter a:

- Una simple normalización por división de cada valor de emisión de cada espectro concatenado por la norma de ese mismo vector X_i . Esta normalización se aplica a cada muestra ($i = 1$ a l muestras), es decir:

$$x_{ij} \text{ (normalizado)} = \frac{x_{ij}}{\|x_i\|}$$

40 donde j es el número de longitudes de ondas del espectro de emisión concatenadas ($j = 1$ a 4545) y $\|x_i\| = \sqrt{\sum_j |x_{ij}|^2}$; o bien

- Una corrección multiplicativa de la dispersión ("Multiplicative scatter correction"), que consiste en realizar una regresión entre el vector x_i (el espectro de emisión concatenado de cada muestra i) y el vector concatenado promedio de todas las muestras ($i = 1$ a l muestras), luego en restar la ordenada en el origen y en dividir por la pendiente respectiva cada intensidad del espectro concatenado de emisión para cada muestra, como se describe a continuación:
- 45

$$x_i = a_i + b_i \bar{x}_i$$

$$x_{ij}(MSC) = \frac{x_{ij} - a_i}{b_i}$$

Véase a propósito de todo esto las referencias: [Bro 1998; Dhanoa 1994].

5 Tras haber tomado y tratado previamente los espectros de fluorescencia de un cierto número de muestras de calibrado, es posible determinar el modelo estadístico multivía que será utilizado para el análisis de otras muestras, a caracterizar. Ello implica el cálculo de los vectores de los factores de carga de ese modelo.

10 Se considera en primer lugar el caso de un modelo multilineal de tipo "PARAFAC" [Bro 1998]. El principio, ilustrado en la figura 5, consiste en descomponer una estructura de tres vías (cubo de datos) X en una suma de productos externos de tres vectores ("tríadas") a_i , b_i y c_i , más un resto E, el cual tiene también forma de un "cubo de datos". Las tres vías que constituyen el "cubo" X son: las muestras, las radiaciones de excitación y las longitudes de onda de emisión. Por lo tanto, se puede escribir:

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk}$$

donde "i" es el índice de las muestras, "j" es el de las radiaciones de excitación, "k" es el de las longitudes de onda de emisión y "f" el de los factores F de descomposición PARAFAC.

15 La concatenación de los espectros permite escribir la descomposición PARAFAC en forma matricial:

$$X_{i \cdot JK} = A(C | \otimes | B)^T$$

donde:

- I, J y K son, respectivamente, los números de muestras, de longitudes de onda de excitación y de longitudes de onda de emisión;
- 20 - $X_{i \cdot JK}$ es la matriz de las fluorescencias concatenadas de todas las muestras;
- B_{JF} y C_{KF} son los vectores (de JF y KJ elementos, respectivamente) de los factores de carga de excitación y de emisión (perfiles bilineales de fluorescencia);
- A_{IF} es el vector (de IF elementos) de las puntuaciones factoriales ("scores") (o intensidades de los factores de carga o "loadings" espectrales);
- 25 - el símbolo $|\otimes|$ representa el producto tensorial de Khatri-Rao;
- el exponente T indica la traspuesta de la matriz columna.

30 Los vectores de los factores de carga se calculan a partir de muestras de calibrado escogidas de forma que sean representativas de la variabilidad total esperada en el seno del grupo de muestras a predecir posteriormente. En otros términos, el modelo es empírico: no es, por tanto, válido más que para muestras similares a las que han servido para el calibrado.

35 Los parámetros A, B y C del modelo PARAFAC se pueden calcular mediante el método de alternancia de los mínimos cuadrados (método iterativo no lineal). En este método se calcula una primera estimación del vector A de manera condicional sobre valores iniciales aleatorios atribuidos a B y C para minimizar la suma de los cuadrados de los restos. A continuación se actualiza el parámetro B utilizando la estimación de A, luego se actualiza el parámetro C utilizando el nuevo valor de B y así sucesivamente. De este modo, cada actualización iterativa de A, B y C mejora la solución (disminución de la superficie de error). El algoritmo converge cuando la mejora de la solución en una iteración se hace muy pequeña (por defecto, este criterio es de 10^{-6}).

40 La convergencia del modelo PARAFAC para datos con poca resolución y que, además, no son verdaderamente trilineales, puede ser problemática [Bro 1998; Harhsman 1984; Rizkallah 2007]. En efecto, la superficie de error del modelo puede contener mínimos locales (puntos más bajos que sus vecinos pero más altos que el verdadero mínimo de la superficie), puntos puerto o cima, terrenos llanos y valles muy estrechos que pueden ralentizar o incluso impedir la convergencia del algoritmo.

45 Los presentes inventores se han dado cuenta de que imponiendo una limitación sobre el número de iteraciones (por ejemplo, un máximo de 30 iteraciones) y/o aumentando el criterio de convergencia (10^{-2} o 10^{-3} en lugar de 10^{-6}) se mejora de manera significativa el modelo en cuanto a los parámetros (factores de carga y puntuaciones factoriales), como se mostrará a continuación con ayuda de la figura 7. En efecto, un número excesivo de iteraciones puede degradar la pertinencia del modelo.

La convergencia se puede facilitar imponiendo una restricción de no negatividad, ya que se sabe a priori que los espectros de fluorescencia así como sus intensidades relativas no pueden tomar valores negativos.

5 Por último, se ha observado que el tratamiento previo de supresión de la difusión de Rayleigh por aplicación de un modelo GLZ facilita la convergencia del procedimiento iterativo de cálculo de los vectores de los factores de carga y de las puntuaciones factoriales del modelo PARAFAC.

En el caso de datos perfectamente multilineales, el modelo PARAFAC no admite en teoría más que una solución única. Sin embargo, en el caso de una MEE limitada a 3 excitaciones, y cuando la fluorescencia se recoge en superficie, la trilinealidad de la matriz MEE está notablemente perturbada.

10 Debido a este hecho, la solución ya no es única y es necesario desarrollar criterios de elección del número de factores, así como del modelo final.

Esta elección está guiada por varios criterios conocidos en sí mismos [Bro 1998; Harhsman 1984]:

- La verificación de la conformidad entre los parámetros espectrales obtenidos (vectores B y C) y el conocimiento a priori de los fluoróforos presentes en la muestra analizada.
- Un criterio conocido con el nombre de CONCORDIA [Bro 1998] que sirve para verificar el porcentaje de desviación del modelo respecto de un modelo multilineal perfecto.
- El estudio del porcentaje de varianza de los datos de fluorescencia explicado por el modelo.
- El estudio de la estructura de los restos, que deben ser aleatorios.
- El estudio de la estructura de las puntuaciones factoriales (vector A) en términos de coherencia de la evolución de las puntuaciones respecto de lo que se espera de la aplicación de la transformación sobre el lote de muestras de calibrado.
- La repetitividad y la reproductibilidad de las puntuaciones etc.

25 Las puntuaciones factoriales ("scores") obtenidas así mediante la descomposición PARAFAC de las MEE de las muestras de calibrado se utilizan a continuación para construir una regresión múltiple (lineal o generalizada) sobre las concentraciones del o de los parámetros de calidad a medir. Los coeficientes de la regresión, así como los vectores B y C se almacenan en una base de datos del modelo.

El modelo se valida por aplicación a nuevas muestras conocidas para las cuales los valores predichos se comparan con los valores medidos mediante los métodos de referencia.

30 Las MEE obtenidas con nuevas muestras se tratan previamente (eliminación de Rayleigh y estandarización) de la misma forma que las MEE de las muestras de calibrado. Se aplican el modelo PARAFAC y el modelo de regresión sobre las nuevas muestras utilizando los coeficientes almacenados (B, C y los coeficientes de regresión obtenidos a partir de las muestras de calibrado. En el caso del empleo de una normalización de tipo MSC, el espectro promedio de las muestras de calibrado proporciona la referencia para corregir los espectros de las nuevas muestras.

A partir del cubo de las matrices tratadas previamente de las nuevas muestras que se van a analizar, se calculan las nuevas puntuaciones factoriales sobre la base de los vectores de los factores de carga B y C como sigue:

$$35 \quad A_{\text{nueva}} = (B \otimes C)^+ * X_{\text{nueva}}$$

donde el exponente + indica la inversa generalizada del producto tensorial y X_{nueva} los datos espectrales concatenados y tratados previamente de las nuevas muestras. Por último, las nuevas puntuaciones se multiplican por los coeficientes de regresión (obtenidos con las muestras de calibrado) para obtener los parámetros de calidad de las nuevas muestras.

40 Así, las operaciones necesarias para predecir las concentraciones de los parámetros de calidad en las muestras nuevas que se van a analizar consisten en:

- tratar previamente los nuevos espectros (eliminar la difusión y estandarizar);
- calcular las puntuaciones factoriales de las muestras que se van a analizar a partir de los vectores B y C memorizados y
- aplicar la ecuación de regresión guardada para obtener las concentraciones de los parámetros.

El método que se acaba de describir ha sido aplicado para predecir el contenido de acrilamida en muestras de achicoria (n = 68).

50 Para cada muestra, se tomaron tres espectros con excitaciones a 280, 340 y 429 nm; se trataron previamente para suprimir la difusión y para estandarización y se concatenaron; luego, las MEE así obtenidas se descompusieron en cuatro factores PARAFAC.

En la figura 7:

- los gráficos de la primera columna, etiquetados "A", representan las puntuaciones factoriales ("scores") de los cuatro factores PARAFAC (respectivamente, puntos negros, círculos grises, cuadrados, estrellas) para 100 muestras;
- los gráficos de la segunda columna, etiquetados "B", representan los factores de carga ("loadings") de excitación de los cuatro factores PARAFAC (curvas F_{1}^{B} , F_{2}^{B} , F_{3}^{B} , F_{4}^{B}) para esas mismas 100 muestras;
- los gráficos de la tercera columna, etiquetados "C" representan los factores de carga de emisión de los cuatro factores PARAFAC (curvas F_{1}^{C} , F_{2}^{C} , F_{3}^{C} , F_{4}^{C}) para esas mismas 100 muestras.

Los gráficos de la primera línea se refieren al caso en el que la difusión de Rayleigh se ha suprimido mediante la introducción de valores perdidos; los de la segunda línea, al caso en el que se ha suprimido mediante la introducción de ceros; los de la tercera y la cuarta líneas, al caso en el que la supresión de la difusión de Rayleigh se ha efectuado utilizando un modelo lineal generalizado.

Con excepción de los datos que incluyen valores perdidos, todos los demás datos fueron también estandarizados mediante MSC.

Los resultados muestran que los valores perdidos en cuanto a la superposición de la difusión y de la fluorescencia (figura 7C línea 1) son mal predichos, mientras que los ceros producen artefactos (figura 7C línea 2) constituidos por variaciones bruscas de intensidad. Los modelos tratados previamente con GLZ producen parámetros espectrales aceptables, con variaciones graduales de intensidad (figura 7C, líneas 3 y 4).

El criterio de convergencia del modelo de la figura 7, líneas 1, 2 y 3 es de 10^{-6} (criterio utilizado comúnmente): el del modelo de la figura 7 línea 4 es de 10^{-2} . La restricción impuesta a la convergencia del modelo ya sea por limitación del número de iteraciones permisible y/o ya sea aumentando el criterio de convergencia mejora de forma muy significativa los resultados del modelo en términos de estructura de las puntuaciones factoriales (figura 7A, líneas 3 y 4) y de separación de los factores (figuras 7B y 7C, líneas 3 y 4). En particular, se puede notar la evolución de la curva F_{2}^{B} que, en la tercera línea de la figura, muestra una contribución de las tres longitudes de onda de excitación; por el contrario, en la cuarta línea, la separación de las excitaciones es netamente mejor. De igual modo, la escarpadura de la curva F_{1}^{C} se reduce considerablemente cuando el factor de convergencia pasa de 10^{-6} a 10^{-2} .

Las puntuaciones factoriales obtenidas con el modelo presentado en la figura 7, línea 4 se utilizaron para calibrar el contenido de acrilamida medido químicamente en las muestras. En la figura 8 se presentan los contenidos de acrilamida medidos químicamente y los predichos con el modelo basado en las puntuaciones factoriales PARAFAC de la fluorescencia medida con el captador. En esta figura, los contenidos de acrilamida medidos de manera convencional se indican en el eje de las abscisas y los predichos por el modelo están en el eje de las ordenadas. N es el número de muestras, R es el coeficiente de correlación y RMSEC es la raíz cuadrada de la media de los cuadrados de los errores. Se notará, en particular, el valor elevado del coeficiente de correlación (casi 95 %); el RMSEC vale 465,47 $\mu\text{g}/\text{kg}$, es decir, aproximadamente 15 % del valor medio del contenido medido de acrilamida.

Cuando la regresión obtenida con las puntuaciones factoriales PARAFAC de fluorescencia no conduce a resultados convenientes (por ejemplo, coeficiente de correlación $R < 0,85$), se aplica otra estrategia dirigida a mejorar la predicción del modelo. Este método se basa en la regresión multivaria NPLS, acoplada preferentemente con una "Corrección Ortogonal de Señal" (OSC, por las siglas en inglés) [Wold 1998].

El tratamiento previo OSC pretende reducir o eliminar la varianza contenida en la señal de fluorescencia que es ortogonal (no correlacionada) a la variable dependiente (y), antes de ejecutar la regresión. El OSC se aplica sobre los datos concatenados después de eliminar la difusión por el método GLZ y después del tratamiento previo de estandarización por MSC.

El algoritmo más frecuentemente utilizado para la corrección OSC es el NIPALS (por sus siglas en inglés de "non linear iterative partial least squares", es decir, algoritmo no lineal e iterativo por mínimos cuadrados parciales). La primera etapa consiste en extraer el vector de las puntuaciones factoriales \mathbf{t} de la primera componente principal (descomposición en valores singulares) y ortogonalizarlo sobre el vector \mathbf{y} : $\mathbf{t}_{\text{nuevo}} = (\mathbf{1} - \mathbf{y}(\mathbf{y}^T\mathbf{y})^{-1}\mathbf{y}^T)\mathbf{t}$. A continuación, se aplica una regresión tipo MLR o PLS (multivariante) para predecir el vector ortogonalizado: $\mathbf{w} = \mathbf{X}^+ \mathbf{t}_{\text{nuevo}}$ (la convergencia se prueba sobre la nueva puntuación factorial $\mathbf{t}^* = \mathbf{X}\mathbf{w}$). Cuando \mathbf{t}^* converge, se calcula el vector propio correspondiente $\mathbf{p} = \mathbf{X}^T \mathbf{t}^*$ y a continuación se elimina el efecto ortogonal de los datos de fluorescencia ($\mathbf{X}_{\text{nuevo}} = \mathbf{X} - \mathbf{t}^*\mathbf{p}^T$). En la aplicación de la invención, se calcula un solo factor OSC y luego se elimina antes de la reorganización de los datos de fluorescencia en una estructura de tres vías. En el caso de una sola variable \mathbf{y} , el modelo converge en una sola iteración.

Al revés que el modelo de descomposición PARAFAC, el modelo de regresión NPLS [Bro 1998] busca factores explicativos a la vez de los datos de fluorescencia \mathbf{X} (tres vías) y de \mathbf{y} (la variable dependiente). La regresión NPLS busca, en consecuencia, una descomposición de los datos \mathbf{X} que conduce a puntuaciones factoriales que tienen una covarianza máxima con \mathbf{y} ; una regresión interna entre las puntuaciones factoriales NPLS de \mathbf{X} e \mathbf{y} conduce a continuación al calibrado:

$$\mathbf{X}^{(I \times J \times K)} = \mathbf{T}(\mathbf{W}^K | \otimes | \mathbf{W}^J)^T + \mathbf{E}_1 ;$$

$$\mathbf{Y} = \mathbf{T}\mathbf{b} + \mathbf{E}_2$$

donde \mathbf{T} es el vector de las puntuaciones factoriales de \mathbf{X} (que tiene la mejor covarianza con \mathbf{y}) y los vectores de los factores de carga respectivos para las excitaciones y las emisiones son \mathbf{W}^K y \mathbf{W}^J .

5 El algoritmo para la regresión NPLS con f factores correspondiente a los datos \mathbf{X} (tres vías) y una sola variable dependiente se describe como sigue:

1. Centrar los datos \mathbf{X} e \mathbf{y} y se pone $\mathbf{y}_0 = \mathbf{y}$
2. Comenzar la iteración $f = 1$
3. Calcular $\mathbf{Z} = \mathbf{X}^T \mathbf{y}$
- 10 4. Determinar \mathbf{w}^J y \mathbf{w}^K descomponiendo \mathbf{z} en valores singulares
5. Calcular los vectores \mathbf{t} y disponerlos en una matriz $\mathbf{T} = [\mathbf{t}_1 \mathbf{t}_2 \mathbf{t}_3 \dots \mathbf{t}_f]$
6. Calcular los coeficientes de regresión $\mathbf{b} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{y}_0$
7. Cada muestra \mathbf{X}_i se reemplaza por $\mathbf{X}_i - \mathbf{t}_i \mathbf{w}^J (\mathbf{w}^K)^T$ e $\mathbf{y} = \mathbf{y}_0 - \mathbf{T}\mathbf{b}$
8. $f = f + 1$. Volver a empezar la fase 1 hasta explicación suficiente de \mathbf{y} .

15 (En estas ecuaciones, el índice del factor f se ha omitido de \mathbf{t} , \mathbf{w}^J , \mathbf{w}^K y \mathbf{b})

Dado que la NPLS se aplica sobre los datos después del tratamiento previo OSC (y por tanto se ha eliminado la mayor varianza perpendicular), solo se utiliza un factor NPLS para la regresión.

20 Después de la concatenación de las MEE de las nuevas muestras, se aplica el tratamiento previo utilizando el espectro promedio de las muestras de calibrado como referencia. Los factores OSC (\mathbf{p} y \mathbf{w}) calculados con las muestras de calibrado se eliminan luego de las nuevas muestras:

$$\mathbf{X}_c = \mathbf{X}_n - \mathbf{X}_n \mathbf{w}^* (\mathbf{p}^T \mathbf{w}^*)^{-1} \mathbf{p}^T ;$$

\mathbf{X}_n es la matriz de las MEE concatenadas y tratadas previamente por MSC y \mathbf{X}_c es la matriz de las MEE concatenadas corregidas por OSC.

25 Los factores de carga (\mathbf{W}^J , \mathbf{W}^K) y los coeficientes de regresión \mathbf{b} de la NPLS obtenidos a partir de las muestras de calibrado se aplican sobre la matriz \mathbf{X}_c para predecir las puntuaciones factoriales de las nuevas muestras y su variable dependiente.

$$\mathbf{T}_n = (\mathbf{W}^J | \otimes | \mathbf{W}^K)^+ \mathbf{X}_c$$

$$\hat{\mathbf{y}} = \mathbf{T}_n \mathbf{b} + \mathbf{e}$$

Se ha aplicado el método NPLS a la predicción del contenido de vitamina C en muestras de judías verdes.

30 Las figuras 9A, 9B, 10A y 10B son gráficos que ilustran la aplicación de un método según un segundo modo de realización de la invención, basado en un modelo NPLS, a la predicción del contenido de vitamina C de diferentes muestras de judías verdes.

35 Tras tomar las matrices MEE de $n = 60$ muestras de judías verdes (figura 9A, espectros en bruto), eliminar la difusión (modelo GLZ) y estandarizar mediante MSC, se aplicó un factor OSC para eliminar de los datos de fluorescencia la mayor parte de la varianza ortogonal (no explicativa) a la variable dependiente (contenido de vitamina C medida químicamente): véase figura 9B espectros con tratamiento previo.

A continuación, se aplicó un modelo NPLS con un factor para calibrar el "cubo" de las MEE tratadas previamente en el contenido de vitamina C medido químicamente.

40 La figura 10A muestra los coeficientes de regresión NPLS para los tres espectros de emisión (excitación a 280 nm: línea negra; excitación a 340 nm: línea gris; excitación a 429 nm: línea discontinua). La figura 10B muestra una gráfica de los contenidos de vitamina C medidos convencionalmente (eje de abscisas) y predichos por aplicación del modelo NPLS a tres espectros de autofluorescencia (eje de las ordenadas). N es el número de muestras, R es el coeficiente de correlación y RMSEC es la raíz cuadrada de la media de los cuadrados de los errores. Se notará, también en este caso, el valor muy alto del coeficiente de correlación (superior a 95 %) y el hecho de que el error cuadrático mediano RMSEC es del orden de 15 %, aceptable para las aplicaciones aquí consideradas.

50 Se ha descrito la invención haciendo referencia a dos ejemplos particulares en los cuales se ha aplicado a la determinación del contenido de una sustancia de interés en muestras alimentarias. De forma más general, la invención puede permitir la determinación de diferentes indicadores de calidad de una muestra; estos indicadores pueden ser de tipo químico, como en los ejemplos previamente citados, pero también de tipo microbiológico (carga microbiana), físico-químico o químico. La invención permite también caracterizar los

procedimientos a los cuales se someten las muestras, proporcionando estimaciones de parámetros tales como el valor de esterilización o de pasteurización, el producto tiempo-temperatura o el valor de cocción de un procedimiento de tratamiento térmico.

5

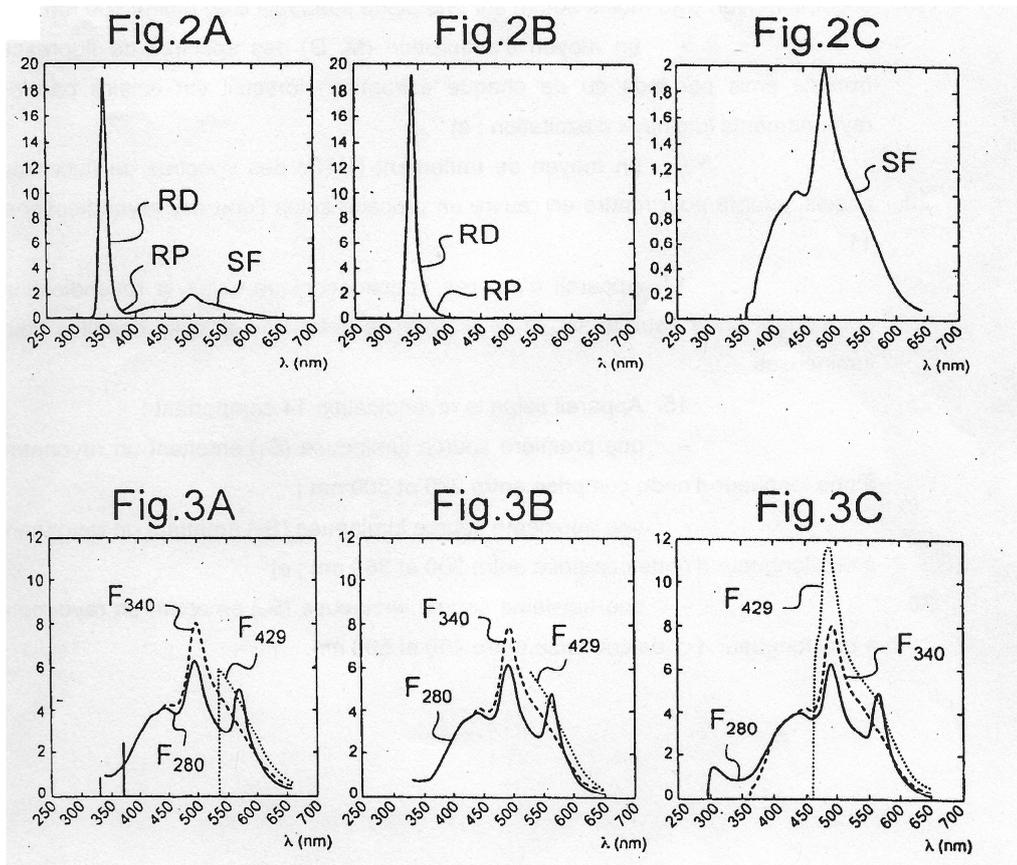
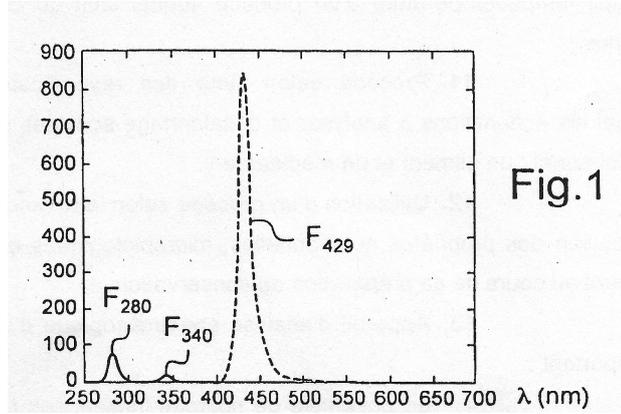
Referencias:

- [Bro 1998] Bro, R. "Multi-way Analysis in the Food Industry Models, Algorithms and Applications" (Análisis multivía en la industria alimentaria. Modelos, algoritmos y aplicaciones); tesis doctoral, Universidad de Amsterdam, 1998
- 10 [Dhanoa 1994] Dhanoa, M.S.; Lister, S.J.; Sanderson, R.; Barnes, R.J.; "The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra", (El enlace entre transformaciones de corrección de dispersión multiplicativa (MSC) y variable normal estándar en espectros en el IR cercano), *J. Near Infrared Spectrosc.*, 1994, 2, 43-47.
- 15 [Harshman 1984] Harshman, R. A., "How can I know if it's real?". A catalog of diagnostics to use with three-mode factor analysis and multidimensional scaling", (¿Cómo se que es real?. Un catálogo de diagnósticos para usar con el análisis de factores trimodal y escalado multidimensional", en *Research methods for multimode data analysis*; editores: H.G. Law, C.W. Snyder, J.A. Hattie, R.P. McDonald, páginas 566-571, Praeger, New York, 1984.
- 20 [Rizkallah 2007] Rizkallah, J., "Chemometric Analysis of Front Face Fluorescence Data – Estimation of Neoformed Contaminants in Processed Foods" (Análisis quimiométrico de datos de fluorescencia frontal – Estimación de contaminantes neoformados en alimentos procesados); tesis doctoral, Agro Paris Tech Francia, 2007.
- 25 [Wold 1998] Wold, S., Antti, H., Lindgren, F., Öhman; "Orthogonal signal correction of near-infrared spectra" (Corrección de señal ortogonal de espectros de infrarrojo cercano), *Chemometr. Intell. Lab. Syst.* 1998, 44, 175-185.
- [Nahorniak 2003] Nahorniak Michelle L., Booksh Karl S.; "Optimizing the implementation of the PARAFAC method for near-real time calibration of excitation-emission fluorescence analysis", (Optimizando el empleo del método PARAFAC para la calibración casi en tiempo real de análisis de fluorescencia de excitación-emisión), *J. Chemometrics*, 2003, 17, 608-617
- 30 [Rizkallah et al., 2008] Rizkallah, J., Lakhali, L., Birlouez-Aragon, I., "Front face fluorescence to monitor food processing and neoformed contamination" (Fluorescencia frontal para controlar el procesado de alimentos y la contaminación por neoformación) en "Optical Methods for Monitoring Fresh and Processed Food – Basics and Applications for a better Understanding of Non-Destructive Sensing", capítulo 4, editor M. Zude, publicado por CRC Press, USA (2008)
- 35 [Davidson 1998] Davidson, R., MacKinnon, J.G., "Econometric theory and methods. Generalized least squares and related methods" (Teoría y métodos econométricos. Mínimos cuadrados generalizados y métodos relacionados), en "Econometric Theory and Methods", Oxford University Press, 2004, páginas 255-261
- 40 [Christensen 2005] Christensen, J., Miquel Becker, E., Fredriksen, C.S. "Fluorescence spectroscopy and PARAFAC in the analysis of yogurt" (Espectroscopía de fluorescencia en el análisis de yogur), *Chemometrics and Intelligent Laboratory Systems*, Elsevier Science Publishers B.V: Amsterdam, Países Bajos, volumen 75, número 2, 28 febrero 2005, páginas 201-208, ISSN: 0169-7439

REIVINDICACIONES

1. Procedimiento de análisis espectroscópico de al menos una muestra que emplea un método de análisis de datos espectroscópicos basado en un modelo estadístico multi-vía, que incluye:
 - 5 a) la iluminación de dicha muestra o de cada una de ellas que se va a analizar (E) mediante una pluralidad de radiaciones luminosas de excitación a sus respectivas longitudes de onda;
 - b) la adquisición de espectros de fluorescencia frontal (F_{280} , F_{340} , F_{429}) de dicha muestra o de cada una de ellas, correspondiente cada uno a una radiación luminosa de excitación respectiva;
 - c) un tratamiento previo de los espectros de fluorescencia adquiridos;
 - 10 d) para cada muestra, el cálculo de un vector de puntuaciones factoriales por aplicación a los espectros tratados previamente de dicho modelo estadístico multivía, identificado por un vector de factores de carga de excitación de la fluorescencia y por un vector de factores de carga de emisión; y
 - e) la determinación de al menos un parámetro escogido entre un indicador de calidad de dicha muestra o de cada una de ellas y un parámetro que caracteriza un procedimiento al cual dicha muestra o cada una de ellas se somete, a partir de dicho vector de puntuaciones factoriales;
 - 15 estando el procedimiento caracterizado por que la separación espectral promedio entre dichas radiaciones luminosas de excitación es de al menos 50 nm, en una zona espectral de al menos 100 nm.
2. Procedimiento según la reivindicación 1, en el cual el número de radiaciones luminosas de excitación y de espectros de fluorescencia correspondientes para cada muestra, está comprendido entre dos y seis, y preferentemente entre 3 y 5, en una zona espectral de al menos 100 nm y, preferentemente, de al menos 150 nm.
- 20 3. Procedimiento según una de las reivindicaciones precedentes, que comprende asimismo una fase preliminar de calibrado que incluye:
 - i) la iluminación de varias muestras de calibrado por dichas radiaciones luminosas de excitación;
 - ii) la toma de los espectros de fluorescencia frontal de las muestras, que corresponden a dichas radiaciones luminosas de excitación;
 - 25 iii) un tratamiento previo de los espectros de fluorescencia tomados;
 - iv) la determinación, mediante un método iterativo, de dichos vectores de los factores de carga de excitación y de emisión del modelo multivía, así como un vector de puntuaciones factoriales ("scores") para cada muestra de calibrado y
 - v) la determinación de una función de regresión que enlaza dichos vectores de las puntuaciones a los valores, conocidos, de dicho o de cada parámetro para dichas muestras de calibrado.
 - 30
4. Procedimiento según una de las reivindicaciones precedentes, en el cual los espectros de fluorescencia son tratados previamente mediante sustracción de una contribución debida a la difusión Rayleigh de primer orden de la radiación luminosa de excitación, siendo calculada dicha contribución por medio de un modelo lineal generalizado.
5. Procedimiento según la reivindicación 4, en el cual dicho modelo lineal generalizado se determina a partir de una primera parte o zona (RD) del espectro a tratar previamente, que comprende solamente dicha contribución debida a la difusión Rayleigh y que se utiliza para predecir la contribución de la difusión Rayleigh en una segunda parte o zona (RP) de dicho espectro en la cual se superpone a una contribución de fluorescencia.
- 35 6. Procedimiento según una de las reivindicaciones precedentes, en el cual dicho modelo estadístico multivía es un modelo PARAFAC, estando representados los espectros de fluorescencia de cada muestra de forma concatenada.
- 40 7. Procedimiento según la reivindicación 6 en el cual los vectores de los factores de carga de dicho modelo PARAFAC se determinan mediante un método iterativo que se detiene cuando la disminución de la función de pérdida debida a la última iteración se hace menor de un valor umbral comprendido entre 10^{-4} y 10^{-2} .
8. Procedimiento según una de las reivindicaciones 1 a 5, en el cual dicho modelo estadístico multivía es un modelo NPLS, estando representados los espectros de fluorescencia de cada muestra en forma concatenada.
- 45 9. Procedimiento según una de las reivindicaciones precedentes en el cual dichas radiaciones luminosas de excitación comprenden:
 - una primera radiación de una longitud de onda comprendida entre 270 y 300 nm;

- una segunda radiación de una longitud de onda comprendida entre 300 y 360 nm; y
 - una tercera radiación de una longitud de onda comprendida entre 400 y 500 nm.
10. Procedimiento según una de las reivindicaciones precedentes en el cual dicho parámetro o al menos uno de los que se van a determinar se escoge entre:
- 5
- el contenido de dicha o de cada muestra en una sustancia de interés;
 - una carga microbiana de dicha o de cada muestra; y
 - un valor de esterilización, de pasteurización o de cocción, o un producto tiempo-temperatura de un procedimiento al que dicha muestra o cada una de ellas se ha sometido.
11. Procedimiento según una de las reivindicaciones precedentes en el cual las muestras que se van a analizar y las de calibrado son muestras de un producto escogido entre un alimento y un medicamento.
- 10
12. Uso de un procedimiento según la reivindicación 11 para medir la evolución de las propiedades nutricionales, microbiológicas y/o toxicológicas de un alimento en el curso de su preparación o conservación.
13. Aparato de análisis espectroscópico de al menos una muestra que incluye:
- 15
- un conjunto de fuentes luminosas (S_1 , S_2 , S_3) para iluminar dicha muestra o cada una de las que se van a analizar (E) mediante radiaciones luminosas de excitación respectivas, que presentan longitudes de onda diferentes con una separación espectral media de al menos 50 nm en una zona espectral de al menos 100 nm;
 - un medio de adquisición (M, D) de los espectros de fluorescencia frontal emitidos por dicha muestra o por cada una de ellas cuando se ilumina por dichas radiaciones luminosas de excitación; y
 - un medio de tratamiento (MTD) de los espectros de fluorescencia tomados, adaptado para llevar a cabo un
- 20
14. Aparato de análisis espectroscópico según la reivindicación 13 que tiene entre dos y seis de dichas fuentes luminosas y, preferentemente, entre tres y cinco.
15. Aparato según la reivindicación 14 que tiene:
- 25
- una primera fuente luminosa (S_1) que emite una radiación de una longitud de onda comprendida entre 270 y 300 nm;
 - una segunda fuente luminosa (S_2) que emite una radiación de una longitud de onda comprendida entre 300 y 360 nm; y
 - una tercera fuente luminosa (S_3) que emite una radiación de una longitud de onda comprendida entre 400 y 500 nm.
- 30



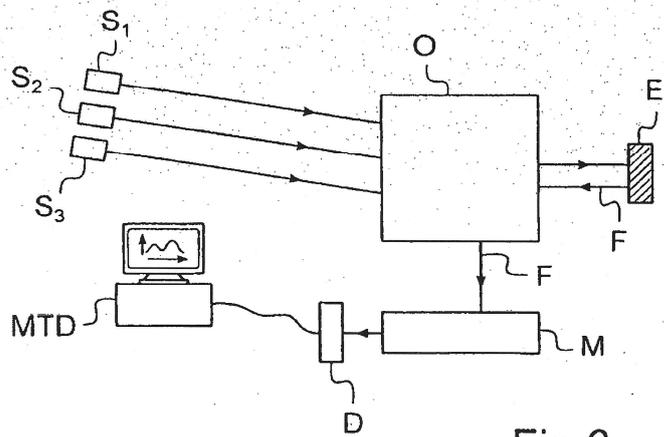
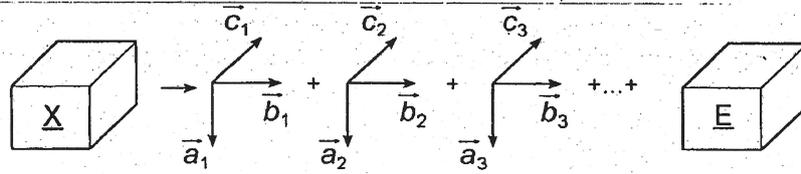
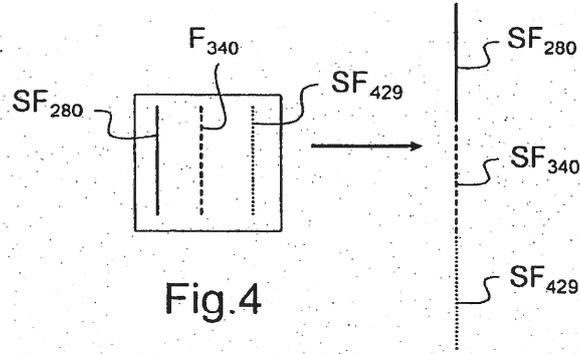


Fig.7

