



Detección de almidón en leche en polvo basado en espectroscopia Raman y mínimos cuadrados parciales

Detection of starch in milk powder based in Raman spectroscopy and partial least squares

José Antonio Blas-Matienzo 

Facultad de Ingeniería en Industrias Alimentarias, Universidad Nacional Agraria de la Selva. Carretera Central km 1.21, Tingo María, Perú.

RESUMEN

Esta investigación tuvo como objetivo establecer un modelo matemático, haciendo uso de información espectral Raman y el algoritmo regresión de mínimos cuadrados parciales (PLS), para pronosticar el porcentaje de adulteración de la leche en polvo por almidón. El modelo de regresión obtenido puede ser usado para identificar muestras que presenten almidón en la leche en polvo en concentraciones en el rango del 5% al 40% (p/p). Se utilizó el método de la validación cruzada con la estrategia de dejar una muestra fuera. El intervalo que resultó óptimo es el rango de número de onda de 2170-2272 cm^{-1} . El modelo de regresión lineal obtenido presenta coeficiente de correlación múltiple de 99,99%, valor mínimo de suma de cuadrados del error residual pronosticado (PRESS) de 237,4 y el valor del estadístico F (19210,29) nos permite establecer que si existe una relación lineal significativa entre las intensidades Raman y los valores de las concentraciones de almidón en la mezcla. El valor del nivel crítico $p=0,006$ indica que, si existe una relación lineal significativa, y, por tanto, que el hiperplano definido por la ecuación de regresión ofrece un buen ajuste.

Palabras clave: mínimos cuadrados parciales; leche en polvo; almidón; espectroscopia Raman; intervalo PLS.

ABSTRACT

This research aimed to establish a mathematical model, using Raman spectral information and the partial least squares regression algorithm (PLS), to predict the percentage of adulteration of powdered milk by starch. The regression model obtained can be used to identify samples that show starch in powdered milk in concentrations ranging from 5% to 40% (w/w). The cross-validation method was used with the strategy of leaving a sample out. The interval that was optimal is the wave number range of 2170-2272 cm^{-1} . The linear regression model obtained has a multiple correlation coefficient of 99.99%, minimum sum of squares of the predicted residual error (PRESS) of 237.4 and the value of the F statistic, 19210.29 allows us to establish that if there is a relationship Linear significance between Raman intensities and the values of starch concentrations in the mixture. The value of the critical level $p = 0.006$ indicates that there is a significant linear relationship, and therefore, that the hyperplane defined by the regression equation offers a good fit.

Keywords: partial least squares; milk powder; starch; Raman spectroscopy; interval PLS.

1. Introducción

El tema de la autenticidad de los alimentos es actual y de gran importancia para investigadores, consumidores, productores e industriales en toda la cadena producción-consumo de leche y sus derivados. Todo producto terminado, e incluso la materia prima, debe cumplir con los requerimientos legales enunciados en su etiquetado, considerando ingredientes, proceso de producción e identidad genética (Escobar, 2014).

Un aditivo no permitido que ha sido agregado a la leche en polvo es el almidón, un polisacárido formado por residuos de α -D-glucosa, compuesto de dos macromoléculas, amilosa y amilopectina. El almidón es agregado para ajustar la densidad y punto de congelación de la leche adulterada. Este producto es ampliamente usado en la industria alimentaria, debido a su bajo valor comercial y fácil acceso (De Almeida *et al.*, 2012).

Generalmente, la identificación es hecha por gel permeación con cromatografía líquida de alta performance (GP-HPLC) (Olieman y Van Den Bedem, 1983). La presencia de almidón en leche es hecha cualitativamente por una reacción colorimétrica entre leche y solución de lugol, donde el desarrollo de color azul indica la presencia de almidón en la leche (De Almeyda *et al.*, 2012).

La espectroscopia infrarroja cercana ha sido extensamente aplicada en el estudio de la leche y sus constituyentes. Iñon *et al.* (2004) evaluaron los componentes principales (carbohidratos, lípidos y proteínas) de diferentes tipos de leche en polvo por espectroscopia infrarroja con Transformada de Fourier (FTIR) con un accesorio de atenuación de la reflexión total (ATR) y calibración multivariada para análisis cuantitativo. Borin *et al.* (2006) propusieron un método para cuantificar el suero, almidón y sacarosa en leche en polvo usando espectroscopia infrarroja cercana (NIR), calibración multivariada y técnicas de inteligencia artificial, a partir de imágenes hiperespectrales.

Sin embargo, a pesar de las numerosas aplicaciones de la espectroscopia Raman, hay pocos estudios de su uso para la calidad de la leche en polvo. Moros *et al.* (2007), evaluaron los parámetros nutricionales de las fórmulas para infantes usando espectroscopia Raman y regresión de mínimos cuadrados parciales (PLS). McGoverin *et al.* (2010) propusieron el uso de la espectroscopia Raman y calibración multivariada (PLS) para la determinación de lípidos y proteínas en la leche en polvo, los resultados obtenidos fueron satisfactorios cuando se compararon con los resultados de la espectroscopia infrarroja.

La espectroscopia Infrarroja y Raman nos provee de información química detallada y se usa rutinariamente en diferentes áreas como la farmacéutica, la de polímeros, forense, ambiental, ciencia de los alimentos, etc. (Rafferty, 2002). A menudo la identificación y cuantificación de componentes en muestras biológicas por métodos espectroscópicos solo se ven obstaculizados por la naturaleza diversa de la muestra. Para obtener información significativa y para una visión más profunda necesitamos procesar y analizar los datos.

Los métodos analíticos que tratan con una sola variable a la vez se denominan métodos univariados. Sin embargo, ninguna de las mediciones espectroscópicas depende de una sola variable. Los datos espectroscópicos

consisten de cientos de variables (número de onda) y mediciones de intensidad Raman en unidades arbitrarias. Para utilizar la información que nos brinda los espectros y manipular el gran conjunto de datos se necesita el análisis multivariado. El análisis de datos multivariado se refiere a los métodos de análisis de datos que tratan con más de una variable a la vez. El principal objetivo de estas técnicas estadísticas de análisis de datos es encontrar la relación existente entre las variables. Esto se basa en la idea de considerar muchas variables no selectivas en lugar de una sola variable y luego finalmente combinarlas en un modelo multivariado. La aplicación de los métodos estadísticos multivariados a la química y a la biología se denominan quimiometría (Deming, 1986). Las herramientas del análisis multivariado se utilizan para efectuar un procesamiento eficiente de la gran cantidad de datos y alinear sus detalles informativos (Svante, 1991). Usando las herramientas analíticas multivariadas, los patrones existentes en los datos, pueden ser modelados y estos modelos pueden ser usados para predecir la adquisición de nuevos datos de un tipo similar. Varios métodos de extracción de datos, tales como, el análisis de componentes principales (PCA), análisis discriminante lineal (LDA), regresión lineal múltiple (MLR), análisis de cluster (CA) y regresión de mínimos cuadrados parciales (PLSR), para nombrar unos pocos son empleados en el campo de la espectroscopia Raman e Infrarroja (Ritz *et al.* 2011). En la presente investigación, haciendo uso de la espectroscopia Raman y la quimiometría, se llegó a establecer una ecuación de regresión que nos permitió pronosticar el porcentaje de adulteración de la leche en polvo por almidón.

2. Material y métodos

2.1. Sistema de espectroscopia Raman

El sistema de espectroscopia Raman fue de la marca Horiba, modelo Xplora Plus Raman Microscope.

2.2. Preparación de la muestra

Para los análisis cuantitativos se prepararon 8 muestras diferentes de la mezcla de leche en polvo con almidón en las concentraciones de 5%, 10%, 15%, 20%, 25%, 30%, 35% y 40% (p/p) de almidón. Para la obtención de los espectros de las muestras estas fueron colocadas en una placa de vidrio. El uso de la placa de vidrio simplifica el experimento y el procesamiento de

datos ya que no interfiere con la señal espectral de la muestra.

2.3. Adquisición espectral

Para desarrollar el método de análisis de cada muestra se obtuvo tres espectros, y los espectros obtenidos fueron agrupados para poder realizar el pretratamiento, que consiste en la obtención del promedio de espectros, el suavizado espectral mediante la aplicación del filtro de Savitzky-Golay y el establecimiento de una línea de base, posteriormente, se realizó la normalización, el cual es un paso previo al desarrollo de un modelo de calibración.

2.4. Análisis espectral y procesamiento de datos

La [Figura 1](#) representa la señal Raman de la leche en polvo sin adulterar. Después de obtener los datos de las intensidades de las señales espectrales Raman, en unidades arbitrarias, para las diferentes mezclas de leche en polvo-almidón en sus diferentes porcentajes, se escogió como rango total, el comprendido en el rango de número de onda de 212-3497 cm^{-1} . Seguidamente se dividió este espectro total en 30 intervalos, conteniendo cada uno de ellos, 30 variables, haciendo un total de 900 variables. Teniendo en cuenta que en la investigación presente no se puede utilizar el centrado-promedio o autoescalado ([Hägglom, 2018](#)), por corresponder esta investigación al tipo de control de calidad, los datos de la intensidad Raman de cada intervalo fueron normalizados, sumando los valores de intensidad Raman de las 30 variables y luego dividiendo el valor de cada variable entre este total, procediendo de este modo para cada porcentaje de la mezcla, y luego sometidos al algoritmo de regresión de mínimos cuadrados parciales, NIPALS (Nonlinear Iterative Partial Least Squares), del programa Minitab 17, procediendo de esta manera para cada intervalo.

3. Resultados y discusión

En la [Figura 2](#) se presentan los espectros Raman obtenidos para las 8 mezclas de leche en polvo-almidón. Todos los modelos son desarrollados sobre la base de los espectros Raman (X) y la variable porcentaje de almidón en la mezcla (y).

Todos los modelos desarrollados son validados por validación cruzada utilizando la estrategia de dejar fuera una muestra, cada vez, para la validación. Esto significa que dado los 8 espectros para calibración y ejecutar el algoritmo de regresión de mínimos cuadrados parciales (PLS), este ejecuta la calibración usando 7 espectros y usa esta calibración para predecir la concentración de la muestra dejada fuera. Este proceso se repite, en este caso, 8 veces hasta que cada espectro haya sido dejado fuera. Los resultados de la aplicación del algoritmo a cada uno de los 30 intervalos se tienen en las tablas 1 y 2.

En la [Tabla 1](#) se resumen los estadísticos del análisis de varianza de los modelos de cada uno de los intervalos y en la [Tabla 2](#) se indican los estadísticos de bondad de ajuste (varianza, error, R^2 , PRESS, R^2 pred) respectivamente. En la [Tabla 2](#) se comparan los resultados obtenidos para cada uno de las ecuaciones o modelos obtenidos al aplicar PLS a cada uno de los 30 intervalos. Según [Haalang y Thomas \(1988\)](#) y [Ni et al. \(2008\)](#) los parámetros, suma de cuadrados del error residual de la predicción (PRESS) y el número de componentes seleccionados se emplean para elegir las regiones espectrales importantes y para eliminar las otras regiones.

El mejor modelo de regresión, basado en los intervalos, debe requerir el menor número de componentes PLS y producir los valores más bajos de PRESS. El Coeficiente de correlación (R^2), Coeficiente de correlación de predicción (R^2 pred) y el valor de la prueba de Fischer, F, y por consiguiente el valor de p, son otros parámetros que ayudan en la selección del mejor intervalo. Teniendo esto en cuenta, el mejor intervalo resultó ser el intervalo número 17, que comprende el rango de número de onda de 2170 - 2272 cm^{-1} .

3.1. Resultados PLSR intervalo

En esta sección se enfoca la situación donde se aplica el algoritmo PLSR a cada uno de los 30 intervalos en los que se dividió el rango total de número de onda 212-3497 cm^{-1} . Cada intervalo consta de 30 variables.

Los valores de intensidad Raman son normalizados según lo indicado anteriormente para cada muestra a fin de poder aplicar el algoritmo

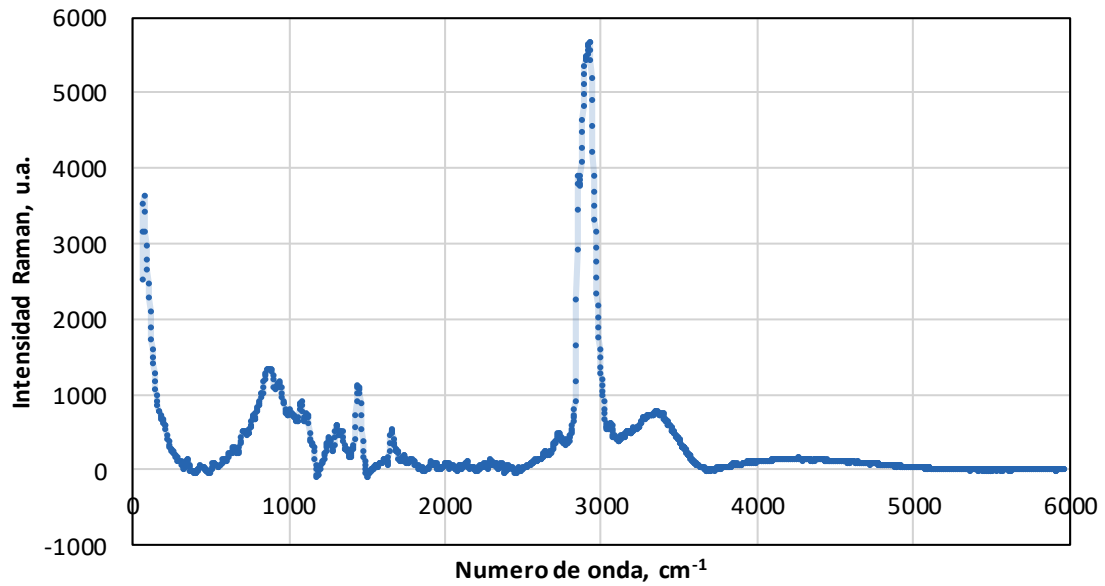


Figura 1. Espectro Raman de leche en polvo sin almidón.

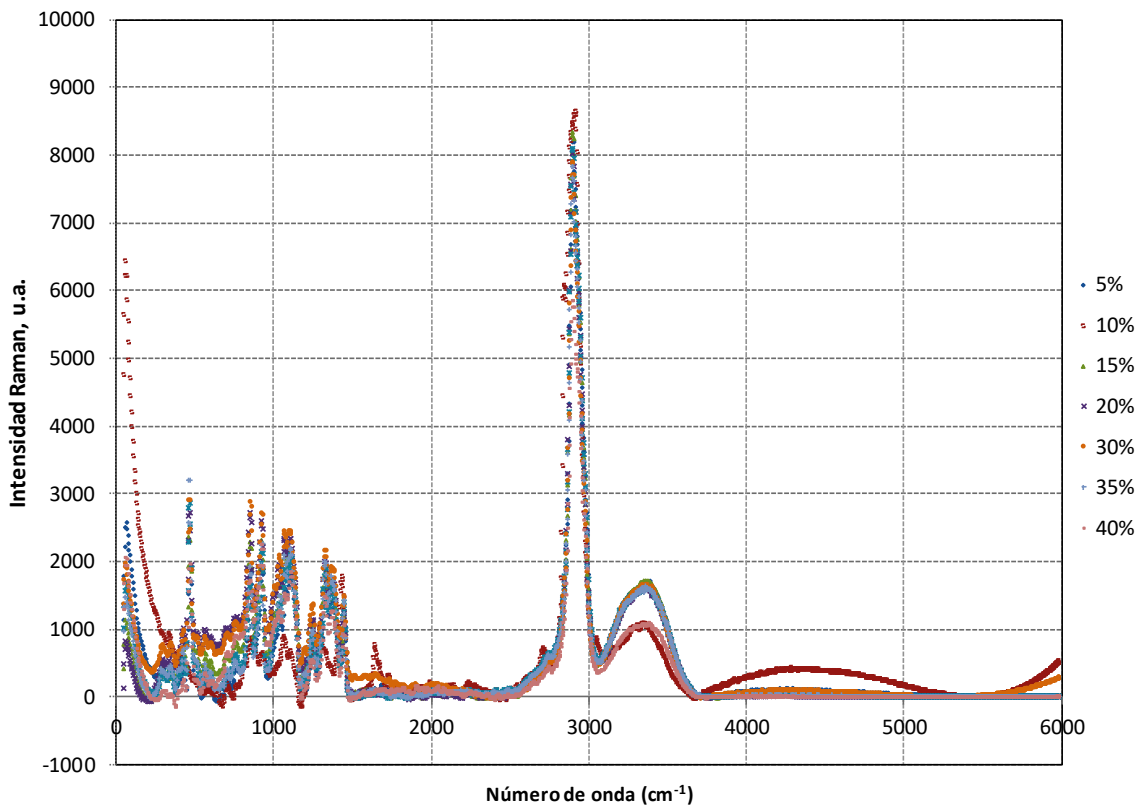


Figura 2. Espectros Raman de las 8 mezclas de almidón con leche en polvo.

Observando los valores resultantes de los parámetros PRESS, R^2 y R^2_{Pred} podemos determinar que el intervalo 17 es el óptimo porque presenta el valor más bajo posible para el parámetro PRESS, los valores más altos tanto para R^2 como para R^2_{Pred} , no existiendo una diferencia abismal entre estos valores, indica que

no hay sobreajuste del modelo y los 6 componentes seleccionados explican 98,98% de la variabilidad en los predictores y la respuesta. En la [Tabla 3](#), se presenta el resumen del análisis de varianza para el intervalo óptimo y en la [Tabla 4](#), un resumen detallado consecuencia de aplicar PLS al intervalo óptimo.

Tabla 1
Resumen de análisis de varianza

N° intervalo	Intervalo, cm ⁻¹	F	P
1	212-347	5,68	0,055
2	351-483	17,28	0,006
3	488-618	11,74	0,019
4	622-750	4,17	0,087
5	754-879	5,81	0,053
6	883-1006	5,83	0,049
7	1011-1132	7,60	0,030
8	1136-1255	1,55	0,260
9	1259-1375	2,93	0,138
10	1379-1494	5,63	0,055
11	1498-1611	13,69	0,014
12	1615-1726	38,54	0,025
13	1730-1839	10,31	0,018
14	1842-1950	0,80	0,404
15	1953-2059	4,02	0,092
16	2063-2166	36,71	0,001
17	2170-2272	19210,29	0,006
18	2275-2376	0,70	0,436
19	2379-2478	9,68	0,021
20	2481-2578	16,90	0,184
21	2582-2677	146,16	0,001
22	2681-2775	24,55	0,003
23	2778-2870	3,15	0,131
24	2874-2965	52,69	0,019
25	2968-3057	3,13	0,150
26	3057-3146	3,10	0,190
27	3149-3236	9,14	0,021
28	3239-3324	7,63	0,030
29	3327-3411	50,18	0,000
30	3414-3497	5,71	0,054

Tabla 2
Estadísticos de bondad de ajuste de los modelos de cada intervalo

N° inter	c.s	Varianza	Error	R ²	PRESS	R ² (pred.)
1	1	0,65799	539,600	0,4860	948,1	0,0970
2	1	0,90620	270,650	0,7422	415,8	0,6040
3	3	0,97805	107,070	0,8980	540,2	0,4855
4	1	0,57660	619,720	0,4097	1782,0	0,0000
5	1	0,50821	533,610	0,4918	1383,8	0,0000
6	2	0,85460	315,179	0,6998	803,6	0,2346
7	2	0,67670	259,837	0,7500	1225,0	0,0000
8	1	0,54060	834,802	0,2049	17431,4	0,0000
9	1	0,85795	705,196	0,3284	1369,6	0,0000
10	1	0,51407	541,720	0,4841	5451,3	0,0000
11	3	0,83749	93,000	0,9112	1331,0	0,0000
12	5	0,95673	10,000	0,9873	1061,3	0,0000
13	1	0,12520	386,000	0,6321	4993,1	0,0000
14	1	0,31709	925,000	0,1182	4476,1	0,0000
15	1	0,41391	629,030	0,4009	2407,2	0,0000
16	2	0,52326	66,952	0,9362	625,0	0,4048
17	6	0,98989	0,009	0,9999	237,4	0,7739
18	1	0,62296	940,893	0,1039	1692,4	0,0000
19	1	0,62011	401,790	0,6173	718,7	0,3155
20	6	0,97977	10,252	0,9902	1040,8	0,0087
21	4	0,80813	5,360	0,9949	698,9	0,3344
22	2	0,77589	97,025	0,9076	2804,5	0,0000
23	2	0,99321	464,995	0,5571	5290,0	0,0000
24	5	0,99873	7,910	0,9925	1288,7	0,0000
25	3	0,93895	314,008	0,7009	6537,8	0,0000
26	4	0,98204	204,424	0,8053	4432,5	0,0000
27	2	0,85865	225,588	0,7852	960,0	0,0857
28	2	0,83371	259,145	0,7532	2083,1	0,0000
29	2	0,69077	49,827	0,9525	929,9	0,1144
30	1	0,65371	538,035	0,4876	902,5	0,1405

c.s.: componentes seleccionados.

Este intervalo presenta el valor más alto de la prueba estadística de Fischer, F, y por consiguiente el valor más bajo de p, 0,006, lo que está de acuerdo con lo que establece Norgaard *et al.* (2000) y Marbach y Heise (1990).

Tabla 3
Resumen del ANOVA para el intervalo óptimo

Fuente	G.L.	S.C.	MS	F	P
Regresión	6	1049,99	174,998	19210,29	0,006
Error Residual	1	0,01	0,009		
Total	7	1050,00			

Tabla 4
Resumen del modelo al aplicar PLS al intervalo óptimo

C	Varianza X	Error	R ²	PRESS	R ² pred
1	0,403342	224,727	0,78597	820,064	0,218986
2	0,667897	48,874	0,95345	354,184	0,662682
3	0,808488	4,711	0,99551	317,330	0,697781
4	0,879915	0,720	0,99931	242,474	0,769072
5	0,963282	0,089	0,99992	242,526	0,769023
6	0,989894	0,009	0,99999	237,377	0,773927

C.: componentes.

Y en la **Tabla 5** se presentan los coeficientes del modelo matemático logrado en el intervalo óptimo.

Tabla 5
Coeficientes del modelo

Número de onda, cm ⁻¹	Coef.	Número de onda, cm ⁻¹	Coef.	Número de onda, cm ⁻¹	Coef.
	40,384				
2170,22	50,287	2205,59	-28,033	2240,76	-2,269
2173,76	90,603	2209,11	-18,960	2244,27	-20,857
2177,31	122,989	2212,64	-45,062	2247,78	-27,336
2180,85	99,308	2216,16	-54,877	2251,28	-37,500
2184,34	245,058	2219,68	9,617	2254,78	-109,278
2187,93	111,146	2223,2	-63,747	2258,28	-50,486
2191,46	135,415	2226,72	-48,789	2261,78	-38,384
2195,00	70,330	2230,30	-204,281	2265,27	-49,234
2198,53	34,240	2233,74	-484,574	2268,77	34,013
2202,06	-22,730	2237,25	48,925	2272,26	31,812

Examinando la **Figura 3** observamos que los puntos siguen una trayectoria lineal indicando que el modelo si se ajusta y predice bien los datos.

Los resultados de esta investigación podrían servir para la fabricación de instrumentos de análisis para la determinación de la adulteración de leche en polvo por almidón cuyos componentes nos proporcionen la radiación Raman solamente en el rango de 2170 - 2271 cm⁻¹, disminuyendo los costos de los instrumentos.

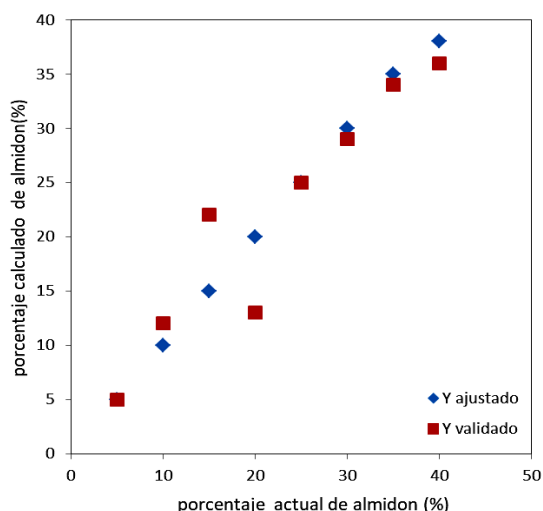


Figura 3. Concentraciones pronosticadas contra concentraciones conocidas.

4. Conclusiones

Se reporta un nuevo método analítico para la determinación cuantitativa de almidón como adulterante en leche en polvo, basado en la utilización de los datos espectrales Raman (X) y la concentración de almidón en la mezcla (y) y la aplicación de la regresión de mínimos cuadrados parciales (PLSR) a cada uno de los 30 intervalos (iPLS) en los que se dividió el espectro total. Resultó como intervalo óptimo el que comprende el rango de número de onda 2120 - 2272 cm^{-1} . La relación entre la intensidad Raman y porcentaje de almidón presenta un coeficiente de correlación de 99,99%. Ante una muestra desconocida de mezcla almidón-leche en polvo, tendrá que encontrarse primero su espectro y los valores de la intensidad Raman en los números de onda y luego serán normalizados y reemplazados en el modelo para determinar el valor de Y. Trabajos futuros podrían usar esta información para desarrollar modelos cuantitativos para la determinación del grado de adulteración de múltiples contaminantes en alimentos sólidos.

Orcid

J. Blas-Matienzo  <https://orcid.org/0000-0002-6160-7143>

Referencias bibliográficas

- Borin, A.; Ferrao M.F.; Mello, C.; Maretto, D.A.; Poppi, R.J. 2006. Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Analytica Chimica Acta* 579(1): 25-32.
- Escobar, A. 2014. Adulteraciones frecuentes en la leche y otros productos lácteos. CONtexto Ganadero. Disponible en: <https://www.contextoganadero.com/blog/adulteraciones-frecuentes-en-la-leche-y-otros-productos-lacteos>
- De Almeida, R.; De Sá Oliveira, K.; Stephani, R.; Cappa de O. 2012. Application of FT-Raman Spectroscopy and Chemometric Analysis for determination of adulteration in milk powder. *Analytical Letters* 45: 2589-2602.
- Deming, S.N. 1986. Chemometrics: an overview. *Clinical Chemistry* 32(9): 1702-1706.
- Hägglblom, K.E. 2018. Basics of Multivariate Modelling and Data Analysis. Disponible en: <http://www.users.abo.fi/khaggblo/MMDA/MMDA6.pdf>
- Haaland, D.; Thomas, E. 1988. Partial Least-Squares Methods for Spectral Analysis 1. Relation to other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Analytical Chemistry* 60(11): 1193-1202.
- Iñon, F.A.; Garrigues, S.; De la Guardia, M. 2004. Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques. *Analytica Chimica Acta* 513(2): 401-412
- Marbach, R.; Heise, H. 1990. Calibration Modeling by Partial Least-Squares and Principal Component Regression and its Optimization Using an Improved Leverage Correction for Prediction Testing. *Chemometrics and Intelligent Laboratory Systems* 9: 45-63.
- McGovern, C.; Clark, S.; Holroy, S.; Gordon K. 2010. Raman spectroscopy quantification of milk powder constituents. *Analytica Chimica Acta* 673: 26-32.
- Moros, J.; Garrigues S.; De la Guardia, M. 2007. Evaluation of nutritional parameters in infant formulas and powdered milk by Raman spectroscopy. *Analytica Chimica Acta* 593: 30-38.
- Ni W.; Brown S.; Man R. Stacked Partial Least Squares Regression Analysis for Spectral Calibration and Prediction. Disponible en: www.interscience.wiley.com.
- Norgaard, L.; Saudland, J.; Wagner, J.; Nielsen, J.; Munch L.; Engelsen S. 2000. Interval Partial Least Squares Regression (iPLS): A comparative Chemometric Study with an Example from Near -Infrared Spectroscopy. *Applied Spectroscopy* 54: 413-419.
- Olieman, C.; Van Den Bedem, J. 1983. A sensitive HPLC method of detecting and estimating rennet whey total solids in skim milk powder. *Neth. Milk Dairy J* 37: 27-36.
- Rafferty, D.; Koenig, J. 2002. FTIR imaging for the characterization of controlled-release drug delivery applications. *J. Control Release* 83: 29-39.
- Ritz, M.; Vaculikova L.; Plevova E. 2011. Application of infrared spectroscopy and chemometric methods for the identification of selected minerals. *Acta Geomater* 8: 47-58.
- Svante, W. 1991. Chemometrics, Why and what and where to next. *J. Pharm Biomed. Anal* 9: 589-96.