



## Método basado en espectroscopía Raman y regresión de mínimos cuadrados parciales para predecir la autenticidad de miel de abeja

Method based on Raman spectroscopy and partial least squares regression to predict the authenticity of bee honey

José Antonio Blas Matienzo\*

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

ORCID del autor:

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

### 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 miel de abeja por jarabe de azúcar. El modelo de regresión obtenido puede ser usado para identificar muestras que presenten jarabe de azúcar en miel de abeja en concentraciones en el rango del 10% al 50% (v/v). 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 643 - 770  $\text{cm}^{-1}$ . El modelo de regresión lineal obtenido presenta coeficiente de correlación múltiple de 99,87%, valor mínimo de suma de cuadrados del error residual pronosticado (PRESS) de 551,26 y el valor del estadístico F, 43,65 nos permite establecer que si existe una relación lineal significativa entre las intensidades Raman y los valores de las concentraciones de jarabe de azúcar en la mezcla. El valor del nivel crítico  $p = 0,022$  indica que, si existe una relación lineal significativa, y, por tanto, que el hiperplano definido por la ecuación de regresión ofrece un buen ajuste. Esta investigación reporta un nuevo método analítico para la determinación cuantitativa del porcentaje de pureza de la miel de abeja adulterado por jarabe de azúcar.

**Palabras clave:** adulteración; espectroscopía Raman; intervalo de mínimos cuadrados parciales; miel de abeja; regresión de mínimos cuadrados parciales.

### 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 bee honey by sugar syrup. The regression model obtained can be used to identify samples that show sugar syrup in bee honey in concentrations ranging from 10% to 40% (v/v). 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 643 - 770  $\text{cm}^{-1}$ . The linear regression model obtained has a multiple correlation coefficient of 99.87%, minimum sum of squares of the predicted residual error (PRESS) of 551.26 and the value of the F statistic, 43.65 allows us to establish that there is a relationship linear significance between Raman intensities and the values of sugar syrup concentrations in the mixture. The value of the critical level  $p = 0.022$  indicates that there is a significant linear relationship, and therefore, that the hyperplane defined by the regression equation offers a good fit. This research reports a new analytical method for the quantitative determination of the purity percentage of honey adulterated by sugar syrup.

**Keywords:** adulteration; bee honey; interval partial least square; Raman spectroscopy; partial least square regression.

## 1. Introducción

La miel es una sustancia dulce natural producida por las abejas a partir del néctar de las flores o de otras partes vivas de la planta y de las secreciones de insectos, que las abejas recogen, transforman, combinan con sustancias específicas propias y almacenan en panales, de los cuales se extrae el producto sin ninguna adición de otras sustancias (Diaz-Forestier et al., 2008).

La miel de abeja es un producto para la salud con alto valor nutricional y es muy popular entre los consumidores. En los últimos años, el consumo interno de miel ha aumentado rápidamente. Sin embargo, la miel se ha convertido en objeto de adulteración con edulcorantes más baratos. La miel adulterada se etiqueta artificialmente como miel pura y el precio de este tipo de sustancias es el mismo que el de la miel pura, que es fraudulenta e injusta para los consumidores. En consecuencia, la discriminación entre la miel no adulterada y la autenticidad de la miel se convierte en un tema muy importante para los procesadores, minoristas y consumidores, así como para las autoridades reguladoras. Existe una demanda creciente de métodos apropiados para garantizar una competencia leal entre los productores y proteger a los consumidores contra el fraude (Chen et al., 2011).

Riswahyuli et al. (2020) utilizaron la espectroscopía infrarroja con transformada de Fourier con reflectancia total atenuada (ATR-FTIR) combinándola con análisis estadístico multivariado para establecer la autenticidad de la miel silvestre de Indonesia. Vasnic et al. (2020) hicieron un enfoque metabolómico asociado con el análisis multivariado y el modelado para discriminar cinco variedades de miel. Se utilizaron técnicas analíticas avanzadas para la determinación de 20 elementos, 14 carbohidratos y la relación de carbono isotópico estable. Wang et al. (2020) trataron la miel de abeja con resinas de adsorción macroporosa y la distinguieron por cromatografía líquida de alta performance con detector electroquímico junto con la quimiometría. Se analizó la caracterización fisicoquímica de la miel de acacia cruda y la tratada con resinas. Además, se establecieron las huellas cromatográficas de las muestras de miel y se analizaron las muestras mediante análisis de componentes principales y proyecciones ortogonales al análisis discriminante de estructuras latentes.

Se caracterizaron y discriminaron tres grupos de miel de acacia rumana, es decir, pura, directa e indirectamente adulterada en función de sus parámetros fisicoquímicos. Se evaluó contenido

de humedad, cenizas, 5-hidroximetilfurfural (HMF), azúcar reductor (fructosa y glucosa) y sacarosa, acidez libre, actividad diastasa, relación entre isótopos de carbono estables de la miel y su proteína. La adulteración condujo a un aumento significativo en el contenido de sacarosa, el nivel de HMF, así como a una disminución en el contenido de azúcar reductor y la actividad diastasa. Se aplicaron análisis de componentes principales (PCA) y análisis discriminario lineal (LDA) a los datos experimentales para distinguir entre miel pura y adulterada (Crăciun et al., 2020). Se et al. (2018) utilizaron la ATR-FTIR junto con la quimiometría para la detección rápida y cuantificación precisa de fructuosa, glucosa, sacarosa, jarabe de maíz y azúcar de caña en miel de abeja (*Heteroglinona itanea*) sin aguijón recolectada en Malasia.

Muestras de miel se adulteraron con jarabes de dátiles y jarabe invertido en tres concentraciones (7%, 15%, 30%). Para cada muestra, se determinaron 32 parámetros que incluyeron índices de color, reológicos, físicos y químicos. Para clasificar las muestras, según el tipo y las concentraciones de adulterante, se aplicó análisis multivariado de componentes principales (PCA) seguido de análisis discriminario lineal (LDA). (Amiry et al., 2017).

La espectroscopía infrarroja y Raman nos proporciona información química detallada que se utiliza de forma rutinaria en diferentes áreas como farmacéutica, polímeros, forense, medioambiental, ciencia de los alimentos, etc. (Rafferty, 2002). A menudo, la identificación y cuantificación de componentes en muestras biológicas mediante métodos espectroscópicos solo se ven obstaculizadas por la naturaleza diversa de la muestra. Para obtener información significativa y 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, los datos espectroscópicos consisten en cientos de variables (número de onda) y mediciones de la intensidad de la señal Raman en unidades arbitrarias. Para utilizar la información proporcionada por los espectros y manipular el gran conjunto de datos, se requiere el análisis multivariado. El análisis de datos multivariados se refiere a métodos que tratan con más de una variable a la vez. El objetivo de estas técnicas estadísticas es encontrar la relación 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.

Las investigaciones anteriormente mencionadas donde aplican PCA y LDA les toma más tiempo establecer la autenticidad de la miel a diferencia del análisis multivariado denominado intervalo de regresión de mínimos cuadrados parciales (iPLSR) que es el que se aplicó en este trabajo de investigación. En la presente investigación, haciendo uso de la espectroscopia Raman y el análisis multivariado se estableció un modelo de regresión que permitió predecir el porcentaje de adulteración de la miel por jarabe de azúcar.

## 2. Materiales y métodos

### a. Sistema de espectroscopia Raman

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

### b. Preparación de la muestra

Para el análisis cuantitativo se utilizó miel de abeja silvestre procedente de la ciudad de Pucallpa y con ella se prepararon 15 muestras de miel de abeja adulterada con jarabe de azúcar en las concentraciones de 0%, 10%, 20%, 30%, 40% and 50% (v/v). Para obtener el espectro de las muestras, estas fueron previamente filtradas y 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.

### c. Adquisición espectral

Para desarrollar el método de análisis de cada muestra se obtuvieron tres espectros, haciendo un total de 45 espectros. Los parámetros experimentales fueron: Laser 638 nm, tiempo de adquisición: 2s, Grating: 600 nm. La [Figura 1](#) representa la señal Raman de la miel de abeja sin adulterar.

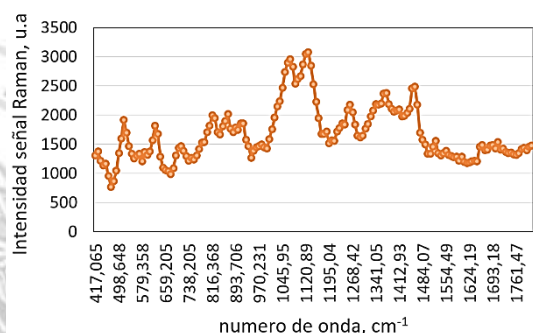


Figura 1. Espectro Raman de la miel de abeja no adulterada.

### d. Preprocesamiento de los espectros

Se utilizó el filtro propuesto por [Savitzky & Golay \(1964\)](#) para obtener el promedio y suavizado de los espectros, el cual se basa en el cálculo de una

regresión (de grado  $k$ ), con al menos  $K+1$  puntos espaciados, para determinar el nuevo valor de cada punto. El resultado es una función similar a los datos de entrada, pero suavizada.

### e. Selección de la región espectral

Se obtuvo los datos numéricos correspondientes a los espectros y se seleccionó la región espectral comprendida entre  $405-1139 \text{ cm}^{-1}$  por presentar picos para aplicar el algoritmo PLSR.

### f. División de la región espectral en intervalos

Basado en el método propuesto por [Norgaard et al. \(2000\)](#) se dividió la región espectral comprendida entre  $405 - 1139 \text{ cm}^{-1}$  en 5 intervalos conteniendo cada intervalo 30 variables.

### g. Normalización

Teniendo en cuenta que en la presente investigación no se puede utilizar el promedio centrado o autoescalado ([Hägglom, 2018](#)), ya que esta investigación corresponde al tipo de control de calidad, se normalizaron los datos de la intensidad Raman de cada intervalo, sumando los valores de la intensidad Raman de las 30 variables y luego dividiendo el valor de cada variable por este total, procediendo de esta manera para cada porcentaje de la mezcla ([Gautan et al., 2015](#)).

### h. Aplicación de PLSR a cada intervalo

Luego fueron sometidos los datos normalizados de cada intervalo al algoritmo de regresión de mínimos cuadrados parciales, NIPALS (Nonlinear Iterative Partial Least Squares) del Programa Minitab 17.

## 3. Resultados y discusión

En la [Figura 2](#) se presenta los espectros Raman obtenidos para las 5 mezclas de miel de abeja – jarabe de azúcar. Todos los modelos fueron desarrollados sobre la base de la intensidad de la señal Raman ( $X$ ) y la variable porcentaje de pureza de la miel de abeja ( $y$ ).

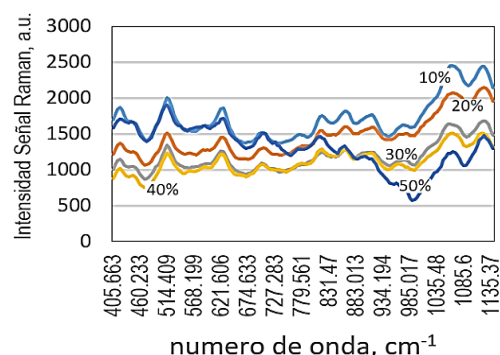


Figura 2. Espectros Raman de las 5 mezclas de miel de abeja con jarabe de azúcar.

**Resultados PLSR intervalo.** En esta sección se enfoca la situación donde se aplica el algoritmo PLSR a cada uno de los 5 intervalos en los que se dividió el rango total de número de onda comprendido entre 405-1139  $\text{cm}^{-1}$ . Para cada intervalo, se construye un modelo PLS. Debido a que cada intervalo espectral puede contener diferente información del analito, tanto como ruido, la precisión de la predicción de cada modelo PLS de cada intervalo es también diferente entre sí. En la aplicación de iPLSR, se seleccionó el modelo PLS con la mejor precisión de predicción.

Todos los modelos desarrollados fueron validados por validación cruzada utilizando la estrategia de dejar fuera, cada vez, una muestra. Esto significa que dado los 5 espectros para calibración y ejecutar el algoritmo de regresión de mínimos cuadrados parciales, este ejecuta la calibración usando 4 espectros y usa esta calibración para predecir la concentración de la muestra dejada fuera. Este proceso se repite, en este caso, 5 veces hasta que cada espectro haya sido dejado fuera.

Se compararon los resultados obtenidos de cada uno de los modelos obtenidos al aplicar PLS a cada intervalo. Según [Haaland & 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 debe requerir el menor número de componentes PLS y producir los valores más bajos de PRESS. El Coeficiente de correlación ( $R^2$ ), el coeficiente de correlación de predicción ( $R^2_{\text{pred}}$ ) 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.

Observando los valores resultantes de los parámetros PRESS,  $R^2$  y  $R^2_{\text{pred}}$  podemos determinar cuál es el intervalo óptimo porque presenta el valor más bajo posible para el parámetro PRESS, los valores más altos tanto para  $R^2$  como para  $R^2_{\text{pred}}$ , no existiendo una diferencia abismal entre estos valores, indica que no hay sobreajuste del modelo y los 3 componentes seleccionados explican 99,87% de la variabilidad en la variable explicativa y la respuesta. Teniendo esto en cuenta el mejor intervalo resultó ser el que comprende el rango de número de onda de 643-770  $\text{cm}^{-1}$ .

En la [Tabla 1](#) se presenta un resumen detallado de los parámetros estadísticos del intervalo óptimo.

**Tabla 1**

Parámetros estadísticos del intervalo óptimo

Componentes	X Varianza	Error	$R^2$	PRESS	$R^2$ (pred.)
1	0,53	208	0,7919	2974,5	0
2	0,93	22,4	0,9776	466,9	0,533
3		1,3	0,9987	551,3	0,448

Este intervalo presenta el valor más alto de la prueba estadística de Fischer, F, y por consiguiente el valor más bajo de p, 0,022, lo que está de acuerdo con lo que establece [Norgaard et al. \(2000\)](#) y [Gaviria et al. \(2016\)](#).

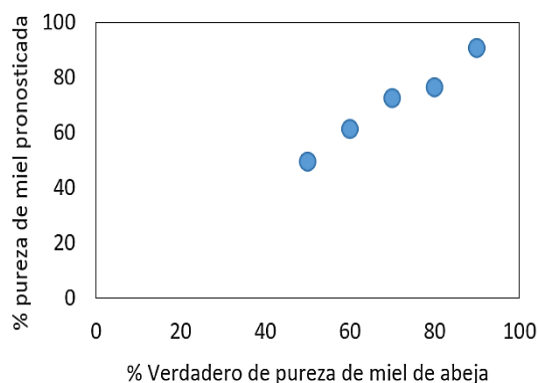
Y en la [Tabla 2](#) se presenta los coeficientes del modelo matemático del intervalo óptimo.

**Tabla 2**

Coeficientes del modelo matemático

Numero de onda, $\text{cm}^{-1}$	Coeficiente	Numero de onda, $\text{cm}^{-1}$	Coeficiente
	788,39		
643,75	331,33	709,77	-1380,24
648,17	116,96	714,16	-2761,31
652,58	15,69	718,53	-3316,42
657,00	-136,73	722,91	-3351,59
661,41	203,14	727,28	-2370,08
665,82	776,38	731,65	-2525,25
670,23	1556,79	736,02	-1690,02
674,63	1490,78	740,39	-407,59
679,04	1228,52	744,75	436,34
683,43	778,90	749,11	714,93
687,83	-1519,08	753,47	491,73
692,22	-1596,01	757,82	359,63
696,62	-3984,53	762,18	283,57
701,00	-2423,55	766,53	323,53
705,39	-2305,79	770,87	238,13

Examinando la [Figura 3](#), donde se representan las concentraciones pronosticadas y las conocidas para evaluar que tan bien el modelo ajusta y predice cada observación, observamos que los puntos siguen una trayectoria lineal indicando que el modelo si se ajusta y predice bien los datos.



**Figura 3.** Porcentajes de pureza de miel abeja pronosticadas contra porcentajes de pureza verdaderas.

Con el algoritmo de regresión mínimos cuadrados parciales se amplía el análisis de componentes principales, con una fase de regresión para los componentes principales de X. La regresión PLS extrae variables latentes para que recojan la mayor variación de las variables X (intensidad señal Raman) de tal manera que se puedan utilizar para modelar las variables respuesta Y (pureza de miel de abeja).

Los resultados de esta investigación podrían servir para la fabricación de instrumentos de análisis para la determinación de la adulteración de miel de abeja por jarabe de azúcar cuyos componentes nos proporcionen la radiación Raman solamente en el rango de 643 - 770  $\text{cm}^{-1}$ , disminuyendo los costos de los instrumentos.

#### 4. Conclusiones

Esta investigación reporta un nuevo método analítico para la determinación cuantitativa del porcentaje de pureza de la miel de abeja adulterado por jarabe de azúcar. Este método está basado en la utilización de los datos espectrales Raman (X) y el porcentaje de pureza de la miel de abeja en la mezcla (y) y la aplicación de la regresión de mínimos cuadrados parciales (PLSR) a cada uno de los 5 intervalos (iPLS) en los que se dividió el espectro total. Resultó como intervalo óptimo el que comprende el rango de número de onda 643 - 770  $\text{cm}^{-1}$ . La relación lineal entre la intensidad Raman y porcentaje de pureza de miel de abeja presenta un coeficiente de correlación de 99,87%. Ante una muestra desconocida de la mezcla miel de abeja-jarabe de azúcar, tendrá que encontrarse primero su espectro y los valores de la intensidad Raman, en los números de onda, indicados en la [Tabla 2](#), y luego recibirán pretratamiento, 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 líquidos.

#### Referencias bibliográficas

- Amiry, S., Esmaili, M., & Alizadeh, M. (2017). Classification of adulterated honey by multivariate analysis. *Food Chem.* 224(1), 390-397.
- Chen, L., Xue, X., Ye, Z., Zhou, J., Chen, F., & Zhao, J. (2011). Determination of chinese honey adulterated with high fructose corn syrup by near infrared spectroscopy. *Food Chem.*, 128, 1110-1114.
- Crăciun, M., Părvulescu O., Donise A., Dobre T., & Stanciu D. (2020). Characterization and classification of Romanian acacia honey based on its physicochemical parameters and chemometrics. *Scientific Reports*, 10, 20690.
- Díaz-Forestier, J., Gomez, M., & Montenegro, G. (2008). Secreción de néctar de quillay. Una herramienta para una apicultura sustentable. *Agronomía y Forestal*, 35, 27-29.
- Gaviria, C., Pérez, R., & Puerta M. (2016). Regresión por Mínimos Cuadrados Parciales con datos de intervalo. *Revista de la Facultad de Ciencias*, 5, 148-159.
- Gautan, R., Vanga, S., Ariese, F., & Umaphathy, S. (2015) Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.*, 2, 8.
- Hägglblom, K. E. (2018). Basics of Multivariate Modelling and Data Analysis. Disponible en <https://www.users.abo.fi/khagglblo/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. *Anal. Chem.*, 60, 1193-1202.
- Ni, W., Brown, S., & Man, R. (2008). Stacked Partial Least Squares Regression Analysis for Spectral Calibration and Prediction. *Journal of Chemometrics*, 23(10), 505-517.
- 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.
- Rafferty, D., & Koenig, J. (2002). FTIR imaging for the characterization of controlled-release drug delivery applications. *J. Control Release*, 83, 29-39.
- Riswahyuli, Y., Rohman, A., Setyabudi, F., & Raharjo S. (2020). Indonesian honey authenticity analysis using attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy combined with multivariate statistical techniques. *Heliyon*, 6(4), e03662.
- Savitzky, A., & Golay, M. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36, 1627-1639.
- Se, K., Ghoshal, S., Wahab, R., Ibrahim, R., & Lani M. (2018). A simple approach for rapid detection and quantification of adulterants in stingless bees (*Heterotrigona itama*) honey. *Food Res Int.*, 105, 453-460
- Vasnic, V., Durdic, S., Tosti, T., Radoiciv, A., Lusic, D., Milojkovic-Opsenica, D., Tesic, Z., & Trifkovic, J. (2020). Two aspects of honeydew honey authenticity: Application of advance analytical method and chemometrics. *Food Chem.*, 305(1), 125457.
- Wang, Q., Zhao, H., Zhang, J., Cheng, N., & Cao, W. (2020). Method for identifying honey adulterated by resin absorption HPLC-ECD coupled with chemometrics. *LWT*, 118, 108863.