

Evaluar el Software Libre R para el Análisis de Espectros Infrarrojo Empleando Quimiometría

Chavarro, Brayan Alexis

Brayan Alexis Chavarro
brayanchavarro249878@correo.itm.edu.co
Instituto Tecnológico Metropolitano, Colombia

Revista Ingenierías USBMed
Universidad de San Buenaventura, Colombia
ISSN-e: 2027-5846
Periodicidad: Semestral
vol. 13, núm. 2, 2022
ingenierias.usbmed@usb.edu.co

Recepción: 12 Febrero 2020
Revisado: 22 Mayo 2021
Aprobación: 22 Julio 2021

URL: <http://portal.amelica.org/ameli/journal/536/5363548005/>

Resumen: En este artículo se usan los datos de espectroscopia infrarroja del artículo de investigación titulado “Data on roasted coffee with specific defects analyzed by infrared-photoacoustic spectroscopy and chemometrics” [1] en el cual realizaron mezclas de granos defectuosos y saludables de café usando las especies *Coffea Arabica* (arábica) y *Coffea canephora* (robusta) en diferentes proporciones. Se usó el software libre R: A Language and Environment for Statistical Computing y el paquete de ChemoSpec para hacer un análisis de componentes principales y un modelo de agrupamiento de los espectros infrarrojo, se logró identificar grupo y tendencias en los datos a través de PCA y se obtuvo un modelo predictivo que logro clasificar las muestras en cinco clases.

Palabras clave: Quimiometría, FTIR, análisis de componentes principales.

Abstract: This article uses infrared spectroscopy data from the research article titled “Data on roasted coffee with specific defects analyzed by infrared-photoacoustic spectroscopy and chemometrics” [1], in which they made mixtures of defective and healthy coffee beans using the *Coffea Arabica* (Arabica) and *Coffea canephora* (Robusta) species in different proportions. The free software R: A Language and Environment for Statistical Computing and ChemoSpec library package were used to make a principal component analysis and a model-based clustering of the infrared spectra data, it was possible to identify groups and trends in the data through PCA and a predictive model was obtained that managed to classify the samples into five classes.

Keywords: Chemometric, FTIR, principal component analysis.

I. INTRODUCCIÓN

Para determinar la presencia y concentración de un compuesto o elemento químico en el medio ambiente se observa por medio de su determinación cualitativa y cuantitativa. La revolución científico-técnica y el desarrollo tecnológico han permitido el desarrollo de diferentes técnicas analíticas. Existen algunas que permiten realizar una evaluación cuantitativa y cualitativa tanto de compuestos inorgánicos como orgánicos [2]. Estas técnicas son muy útiles debido a su facilidad para cuantificar, porque no destruyen la muestra — como puede suceder con los solventes—, reducen el tiempo para la cuantificación, reduce el uso de solventes que pueden ser perjudiciales tanto para la salud humana como el medio ambiente y, por lo tanto, se reducen muchos los gastos económicos [3].

Existen diferentes compuestos —tanto orgánicos como inorgánicos— que son importantes caracterizar, dado que son fuentes de muchos productos que son esenciales para el hombre: uno de ellos son los taninos.

Los taninos son compuestos polifenólicos solubles en agua formados por el metabolismo secundario en las plantas, se pueden encontrar en varias partes de la planta, como raíces, ramas, flores, hojas, frutos y semillas.

Estudios epidemiológicos han demostrado que el consumo de frutas y hortalizas frescas (FHF) juega un papel importante en la humanidad, debido a los efectos positivos en la prevención de enfermedades crónicas degenerativas como las cardiovasculares, distintos tipos de cáncer y problemas neurológicos [4]. Este efecto protector ha sido atribuido a los compuestos bioactivos, conocidos como compuestos fenólicos, encontrados de manera natural en las FHF; además, la mayoría de ellos poseen actividad antioxidante [5].

Para la cuantificación de estos polifenoles, se utiliza un método tradicional llamado ensayo de FolinCiocalteu. Aunque este método aún sigue siendo empleado y aceptado por la comunidad científica, presenta algunos inconvenientes: un estudio muestra cómo este método puede ser inespecífico e interactuar con otras moléculas como azúcares, que están presentes en las hortalizas, lo que puede alterar los resultados encontrados [6], donde observaron que la presencia de fructosa produce un efecto sinérgico con el ácido gálico al reaccionar con el reactivo Folin-Ciocalteu reflejado en un aumento de la señal. Así, al tener concentraciones elevadas de fructosa o sustancias reductoras, puede sobreestimar la cuantificación de compuestos polifenólicos dando un resultado erróneo. Sin embargo, gracias a los avances tecnológicos como la espectroscopia infrarroja con transformada de Fourier (FTIR), una de las técnicas más rápidas y limpias, pues no se requieren reactivos ni pretratamiento de las muestras, ayuda a disminuir el tiempo para la cuantificación y puede mejorar la eficiencia de cuantificación comparada con el método de Folin-Ciocalteu al proporcionar información sobre las principales bandas vibratorias [7]. Esta técnica es uno de los métodos de análisis más usados y su principio de funcionamiento se basa en el empleo de las interacciones entre la radiación electromagnética y la materia [3].

Pero para tener un resultado aún más eficiente y correcto el FTIR se combina con la quimiometría para tratar, específicamente, datos complejos de técnicas analíticas que pueden ser correlacionados con características de la muestra. La quimiometría utiliza métodos de origen matemático, estadístico y otros procedentes del campo de la lógica formal para conseguir sus fines [8].

A continuación, se presentan algunos ejemplos del uso de la quimiometría. Salazar *et al.* [9] emplearon técnicas quimiométricas para el control de calidad de la Peroxisomicina A1 (PA1), tanto en CLAR como en espectrofotometría UV-visible. El uso de cromatogramas derivados en CLAR permitió resolver la señal de un contaminante con tiempo de retención cercano al de la PA1 y de esta manera mejoraron los parámetros de cuantificación de este. Por UV-visible se realizó la cuantificación del contaminante empleando dos técnicas quimiométricas: la espectroscopía de derivadas y la calibración multivariante. Ambos métodos fueron comparados en función de precisión y sensibilidad.

Por otro lado, Rohman *et al.* [10] hicieron un estudio con espectroscopia FTIR combinada con quimiometría para el análisis de adulteración con grasa de cerdo de aceites vegetales donde realizaron el análisis quimiométrico de mínimos cuadrados parciales (PLS) y análisis discriminantes (DA) en la región de 1500–1000 cm^{-1} que se usaron para cuantificar y clasificar la grasa de cerdo en la mezcla con aceites vegetales, respectivamente. La calibración PLS fue usada satisfactoriamente para la cuantificación de grasa de cerdo en la mezcla con aceites vegetales, bien sea usando un espectro normal o sus primeras derivadas. Además, DA basados en la distancia Mahalanobis pueden clasificar la grasa de cerdo en aceites vegetales.

Stöbener *et al.* [11] realizaron un estudio de la viabilidad de la espectroscopia de infrarrojos por transformada de Fourier de reflectancia total atenuada (ATR FTIR) y la modelización quimiométrica para el análisis de compuestos traza, en el ejemplo del furfural (hidroximetil) en la miel hasta una concentración de 110 ppm; se comparan los resultados con los de la literatura. Realizaron mediciones repetidas aleatorias de las muestras de calibración para evitar artefactos analíticos. Calcularon un modelo de mínimos cuadrados parciales (PLS) para la determinación cuantitativa. Finalmente, indican que es posible la determinación de

compuestos de trazas con la espectroscopia FTIR ATR, aunque los esfuerzos de calibración y modelización son mayores en comparación con las aplicaciones estándar de la espectroscopia FTIR ATR y la quimiometría.

Christou *et al.* [12] analizaron muestras de algarrobo de siete países mediterráneos diferentes (Chipre, España, Grecia, Italia, Turquía, Jordania y Palestina) mediante espectroscopia de infrarrojos por transformada de Fourier (FTIR). Examinaron muestras de semillas y carne de cultivares autóctonos y extranjeros, tanto auténticos como comerciales. Usaron técnicas quimiométricas multivariadas, entre ellas el análisis de componentes principales (PCA), el análisis de conglomerados (CA), los mínimos cuadrados parciales (PLS) y el análisis discriminante de mínimos cuadrados parciales ortogonales (OPLS-DA). Concretamente, el PCA no supervisado enmarcaba la importancia de la variedad de algarrobas, mientras que el análisis supervisado destacaba la contribución del origen geográfico. Los mejores modelos de clasificación se lograron con la regresión del PLS en los espectros de la primera derivada, lo que dio una clasificación general correcta. Así pues, la metodología aplicada permitió diferenciar la carne y la semilla de las algarrobas según su origen. En conclusión, exponen que este método es una herramienta rápida y poderosa para la discriminación exitosa del origen y tipo de algarrobas [12].

Por lo tanto, dada la utilidad de la quimiometría, en este artículo se quiere hacer una comparación de datos experimentales tomados del artículo de Dias *et al.* [13], pero procesando las señales empleando software libre.

II. METODOLOGÍA

A. Datos de FTIR

Los datos se obtuvieron de la revista Data In Brief del artículo científico titulado “Data on roasted coffee with specific defects analyzed by infrared-photoacoustic spectroscopy and chemometrics”; los datos mantuvieron el formato original y fueron procesados mediante el programa R usando Rstudio [1].

Para el desarrollo de la metodología los autores del artículo hicieron lo siguiente:

Las muestras de granos saludables de *Coffea Arábica* (arábica) y *Coffea canephora* (robusta) y 25 muestras de granos defectuosos y saludables de café Arábica, fueron suministradas por el Instituto Agronómico do Paraná (IAPAR).

Los selectores de café capacitados de IAPAR clasificaron manualmente cada selección grano por grano. Las selecciones contenían granos de café rotos, ácidos, negros y saludables, piel y maderas de café.

Además de los granos sanos y enteros ya incluidos en cada selección, cada una se mezcló adicionalmente en una base de granos de café saludables en las proporciones del 20 % y el 40 % (p/p) donde utilizaron tres de estas bases: primero 100 % de café arábica, segundo una mezcla de arábica a robusta en las proporciones de 80:20 y tercero una mezcla de arábica a robusta en las proporciones de 50:50 (p/w). Así, considerando que cada una de las tres bases se mezcló con cada selección al 20 % y al 40 % de selección en la muestra final, esta ascendió a 150 muestras compuestas por selecciones y bases. Sumando al análisis, las tres bases puras y una muestra de café robusta saludable se analizaron 154 muestras.

Todas las muestras fueron tostadas en grado medio usando un equipo Probat Emmerich am Rhein, Alemania, modelo PRG1Z, ERD Gas, correspondiente al 17 % de pérdida de peso, y entre 22 y 26 de luminosidad (colorímetro portátil Konica Minolta BC-10). El café se molió en un molinillo Ditting, modelo KR805.

Los autores [1] tomaron los espectros FTIR usando un espectrómetro BRUKER FTIR modelo Tensor 37 acoplado a un detector foto-acústico Gasera modelo PA 301, todas las muestras se midieron por triplicado y obtuvieron el espectro promedio.

En una celda circular de metal PAS de 9 mm de diámetro y 5 mm de profundidad, que contenía la muestra de café RG, se aisló herméticamente de la atmósfera de la habitación y se purgó con helio durante 1 minuto

antes del análisis para reducir el vapor de agua y el dióxido de carbono en la cámara de muestra. La luz infrarroja se enfocó en la muestra y se detectó la señal PAS. El espectro PAS se obtuvo de la media de 16 exploraciones con una resolución de 4 cm^{-1} en una región de número de onda de 600 a 4000 cm^{-1} . El tiempo total de medición por muestra no fue superior a 2 minutos. Antes del análisis, la señal PAS se calibró con una muestra estándar de polietileno.

B. Análisis de datos

Los 154 espectros promedio obtenidos por FTIR y sus valores de intensidad a 1763 valores de longitud de onda se organizaron como una matriz de 154×1763 , donde se usó PCA (análisis de componentes principales) mediante el software libre R: A Language and Environment for Statistical Computing, Vienna, Austria, 2020, R, y el paquete de quimiometría ChemoSpec.

El análisis de componentes principales (PCA) es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones, a la vez que conserva su información [14]. El método PCA pertenece a la familia de técnicas conocida como Unsupervised Learning. Los métodos de Supervised Learning tienen el objetivo de predecir una variable respuesta Y a partir de una serie de predictores. En el caso de Unsupervised Learning, la variable respuesta Y no se tiene en cuenta, ya que el objetivo no es predecir Y sino extraer información empleando los predictores, por ejemplo, para identificar subgrupos [15]. El principal problema al que se enfrentan los métodos de Unsupervised Learning es la dificultad para validar los resultados, dado que no se dispone de una variable respuesta que permita contrastarlos [16].

En la Tabla 1 está la nomenclatura usada para dichas muestras.

TABLA 1.
Nomenclatura de las muestras de café

Nombre	Muestra
20A	Arábica y 20 % de selección con defectos
40A	Arábica y 40 % de selección con defectos
20AR20	Mezcla Arábica 80 % y Robusta 20 % y 20 % selección con defectos
40AR20	Mezcla Arábica 80 % y Robusta 20 % y 40 % selección con defectos
20AR50	Mezcla Arábica y Robusta 50 % y 20 % selección con defectos
40AR50	Mezcla Arábica y Robusta 50 % y 40 % selección con defectos
A100	Arábica
AR20	Arábica 80 % y Robusta 20 %
AR50	Arábica y Robusta 50 %
R100	Robusta 100 %

III. RESULTADOS Y DISCUSIÓN

A. Espectros FTIR

A partir de los espectros adquiridos, se generó una matriz de datos de 154 líneas (muestras) y 850 columnas. Sobre esta matriz, se realizó un análisis exploratorio mediante PCA y funciones quimiométricas de Rstudio.

En la Figura 1 se aprecian los espectros FTIR característicos de cada muestra.

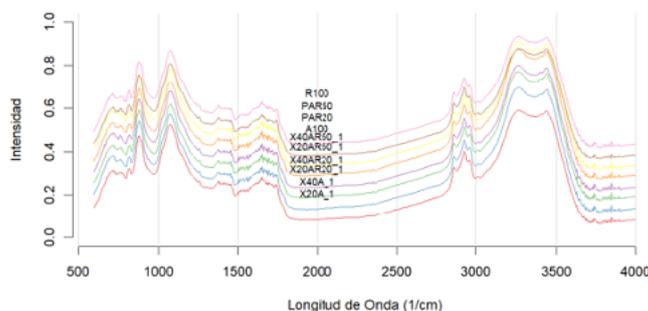


FIGURA 1.
Espectros FTIR de las mezclas de café

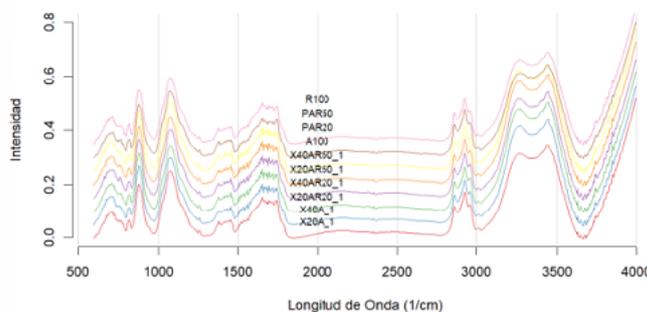


FIGURA 2.
Espectros FTIR de las mezclas de café con corrección de línea base y remoción de longitudes por encima de 3700 cm^{-1}

Previo al análisis de los espectros FTIR se realizó una corrección de la línea base, empleando `BaselineSpectra` del programa R y se eliminaron las frecuencias por encima de 3700 cm^{-1} . Así, se obtiene la Figura 2.

B. Análisis PCA

El primer análisis que se realizó fue un jerárquico de clúster para agrupar los elementos tratando de lograr la máxima homogeneidad dentro de cada grupo y la mayor diferencia entre los distintos grupos. En la Figura 3 se aprecia que existen 2 grupos principales, que a su vez se subdividen en otros grupos.

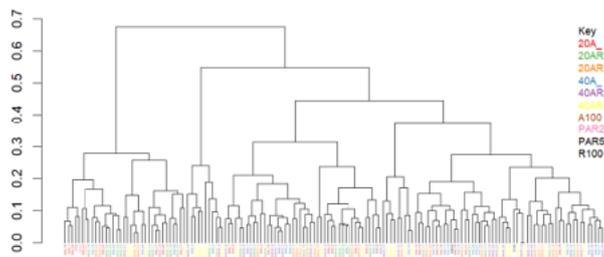


FIGURA 3.
Análisis jerárquico de clusters

Se realizó un análisis de componentes principales para identificar el número de componentes necesarios para el análisis. Para ello, se usó la función `Scree` de `ChemoSpec`; los resultados se muestran en la Figura 4,

en donde los cuatro primeros componentes principales (PC1, PC2, PC3 y PC4) logran explicar el 95 % de la varianza.

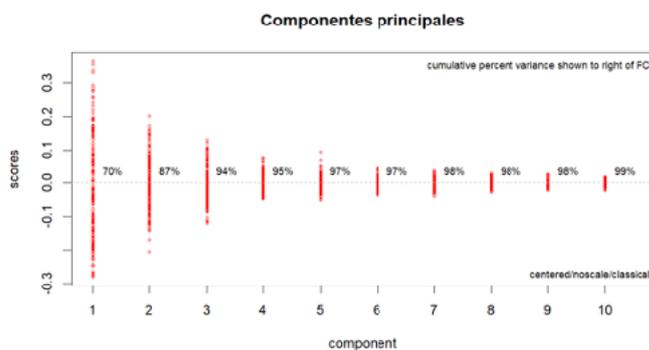


FIGURA 4.
Componentes necesarios

El análisis de las puntuaciones (scores) del análisis de componentes para PC1 vs. PC2, se muestra en la Figura 5. Se observan dos discriminaciones de grupos para PC1, lo que logra explicar el 70 % de la varianza.

El mismo análisis se realizó para las puntuaciones de PC1 y PC3 que se muestran en la Figura 6. En este caso no se observan diferencias, para las puntuaciones de PC1 y PC4, expuestas en la Figura 7, nuevamente se observan dos zonas discriminadas en PC1.

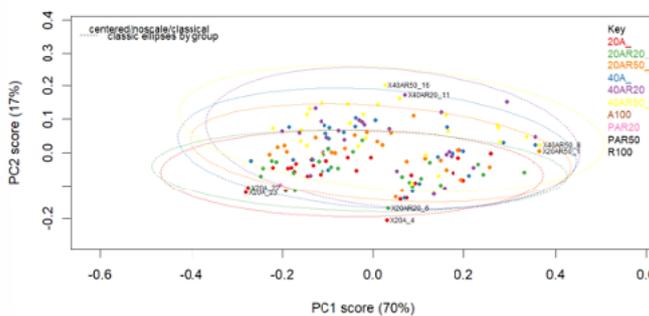


FIGURA 5.
PCA PC1 vs. PC2

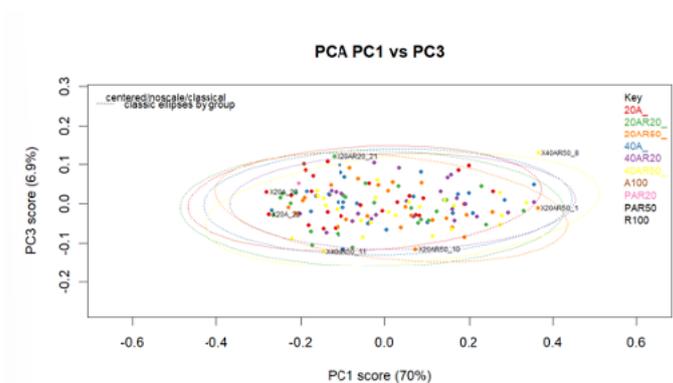


FIGURA 6.
PCA PC1 vs. PC3

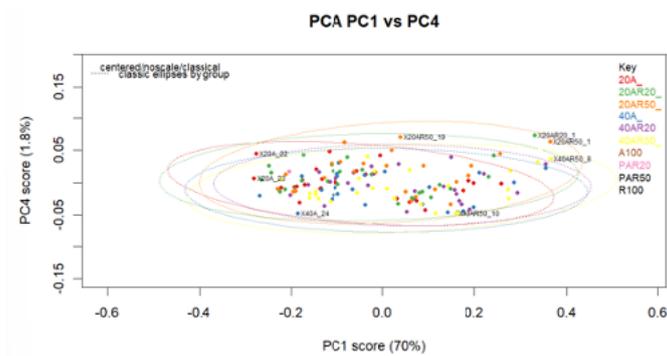


FIGURA 7.
PCA PC1 vs. PC4

C. Agrupamiento basado en modelos

Este modelo asume que los datos están agrupados y que están normalmente distribuidos. Para ello, se usa la librería Mclust dentro del paquete ChemoSpec; lo primero que se hace es buscar el modelo que mejor ajuste la distribución de los datos usando el criterio de información Bayesiano (BIC). Al aplicar esto a los datos de FTIR del café obtenemos la Figura 8 donde se aprecia que el mejor modelo es el EEE, donde se muestran agrupamientos elipsoides con igual forma, volumen y orientación [17].

Empleando el modelo EEE, Mclust encuentra el siguiente agrupamiento, mostrado en la Figura 9. Se

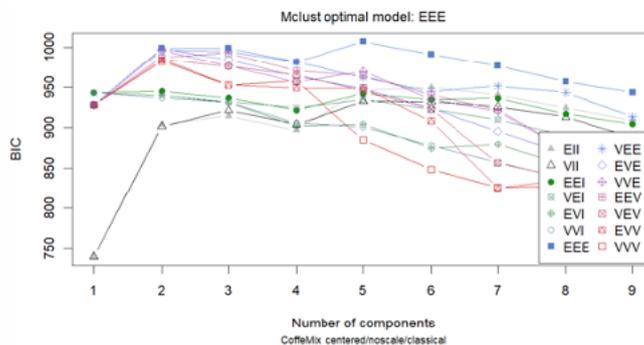


FIGURA 8.
Ajuste de datos criterio BIC

aprecian solamente 2 agrupamientos, pero las muestras consisten en 3 tipos de café y 2 tipos de modificación con una selección (20 % y 40 %). Esto no se evidencia en los resultados, es por ello que se realizó un análisis para seleccionar las frecuencias del espectro FTIR que poseen mayor influencia con el fin de hacer nuevamente el análisis.

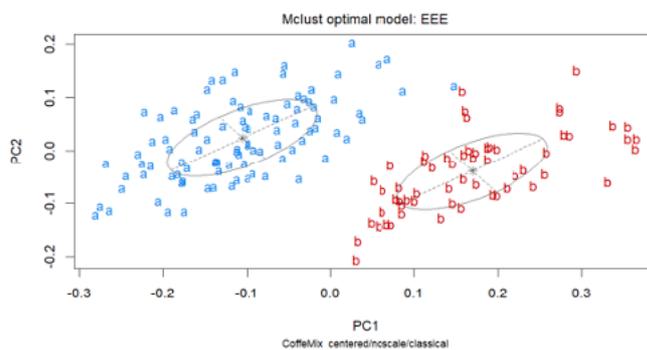


FIGURA 9.
Grupos encontrados en los datos

Para buscar las frecuencias que más influencia tienen sobre la varianza, se realizó un análisis s-Plot sobre los datos, Figura 10, claramente se observa que las bandas alrededor de 800, 2000 y 3000 cm^{-1} son las que más influencia poseen, por esto se realizó nuevamente el análisis a partir de los datos, pero removiendo las frecuencias comprendidas entre 2000 y 3000 cm^{-1} .

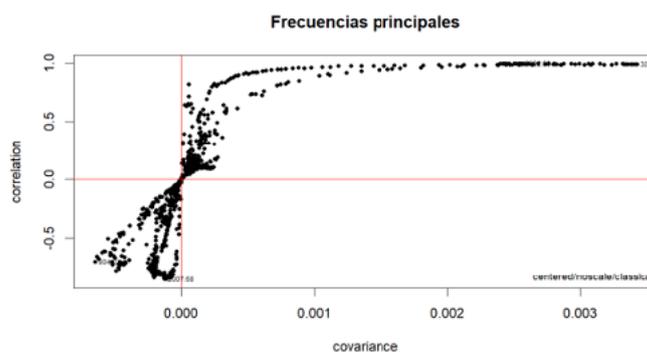


FIGURA 10.
Identificación frecuencias importantes

D. Análisis restringiendo frecuencias

En la Figura 11 se aprecia el análisis jerárquico de clusters, cuyo resultado es similar al anterior: se tienen 2 grupos principales, pero las subdivisiones de estos cambian, por lo que nuevamente se realiza un análisis de agrupamiento basado en modelos, de donde se concluye nuevamente que se ajusta al modelo EEE, mostrado en la Figura 13.

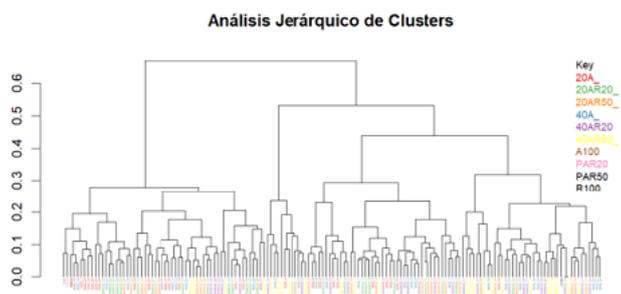


FIGURA 11. Análisis jerárquico de clusters frecuencias importantes

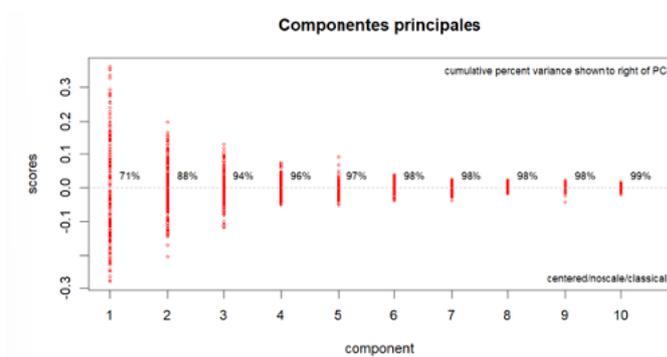


FIGURA 12. Componentes necesarios frecuencias importantes

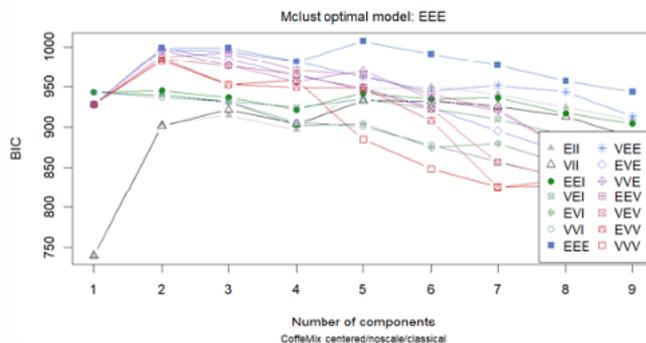


FIGURA 13. Ajuste de datos criterio BIC frecuencias importantes

Después de hacer el análisis de agrupamientos con Mclust para estos nuevos datos (Figura 14), se encuentran que ya existen 5 grupos donde 2 de ellos están claramente diferenciados (c y b) y 3 de ellos se superponen (a, e y d), esto podría corresponderse con los tres tipos de café empleados y las dos modificaciones por medio de una selección al 20 % y 40 %.

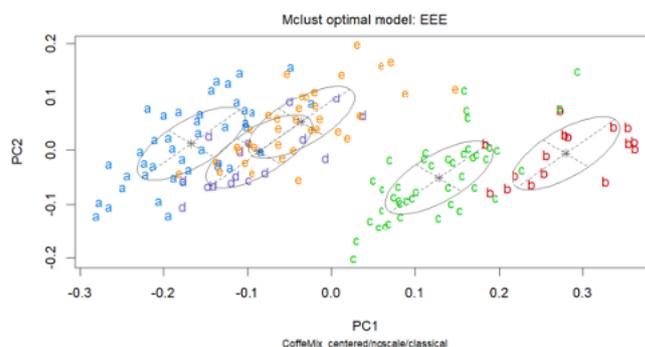


FIGURA 14.

Grupos encontrados en los datos frecuencias importantes

E. ANOVA-PCA

Se puede hacer un análisis de varianza (ANOVA) combinado con PCA, para ello se debe subdividir la matriz de datos en los factores que se usaron para el experimento.

En el artículo original [1], emplearon 2 factores:

Factor 1: Porcentaje de selección usado para modificar el café. 20 % y 40 %. Se asignan los códigos a para 20 % y A para 40 %.

Factor 2: Tipo de base de café empleado, arábica, mezcla Arábica 80 % y robusta 20 % y mezcla arábica robusta 50 %. Se asignan los códigos b para Arábica, c para mezcla Arábica 80 % y Robusta 20 % y d para arábica aobusta 50 %.

En la Tabla 2, se muestra la codificación para los factores mencionados.

TABLA 2.
Factores para el ANOVA-PCA

Nombre	Codificación
20A	ab
40A	Ab
20AR20	ac
40AR20	Ac
20AR50	ad
40AR50	Ad

Para facilitar el análisis se van a colorear las muestras para ver si las diferencias son por el tipo de factor (ver Tabla 3).

TABLA 3.
Colores para los factores

Código	Color
a	Negro
A	Naranja
b	Rojo
c	Azul
d	Verde

En la Figura 15 se ve el análisis para el factor 1. En la parte A (color por factor 2) y en la parte B (color por factor 1) se observa una división en 2 grupos por PC1 que no está relacionada con ninguno de los factores; respecto a PC2, en la Figura 15B se observa una separación de los puntos por color, indicando que podría existir una diferencia entre los espectros por el porcentaje de selección en la muestra de café.

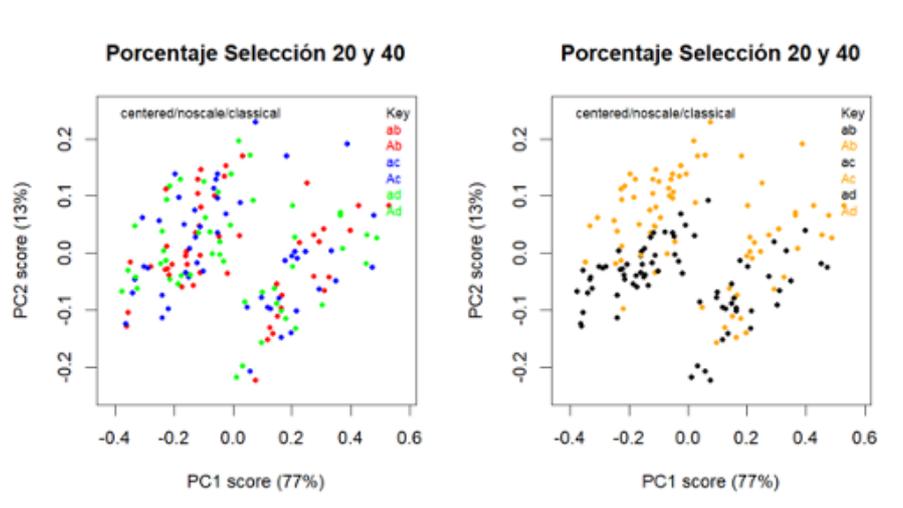


FIGURA 15.
ANOVA-PCA Factor 1

En la Figura 16 se muestra ahora el análisis para el factor 2. En la parte A (color por factor 2) y en la parte B (color por factor 1), nuevamente parecen existir dos grupos por PC1 no relacionados con los 2 factores, respecto a PC2 no se aprecian cambios en los datos debido al tipo de base de café empleado.

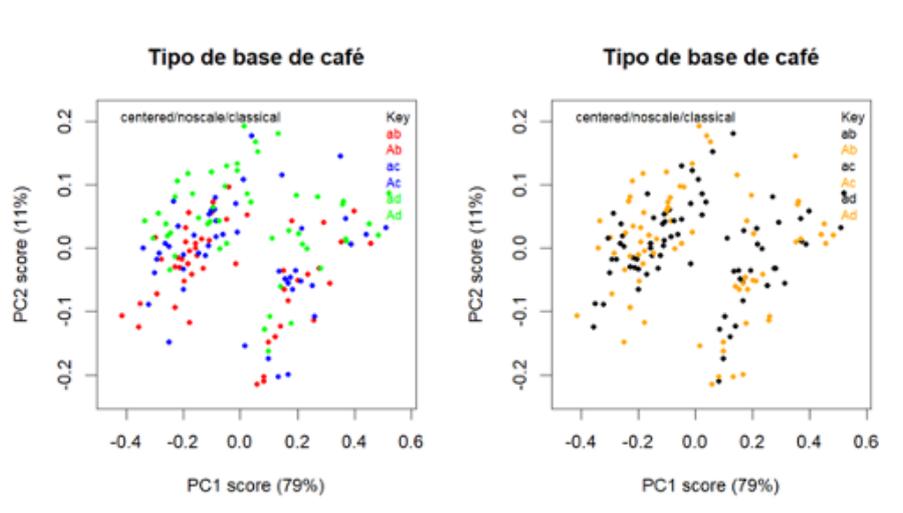


FIGURA 16.
ANOVA-PCA Factor 2

IV. CONCLUSIONES

El análisis de los espectros FTIR de muestras de café, empleando PCA, así como el agrupamiento basado en modelos, permitió identificar grupos en los datos, los que pueden estar relacionados con los factores experimentales empleados de porcentaje de modificación y tipo de base de café.

El software gratuito R junto con las librerías de quimiometría ChemoSpec y de agrupamiento basado en modelos Mclust es una herramienta poderosa para el análisis de espectros FTIR de muestras complejas.

V. AGRADECIMIENTOS

Los autores agradecen el apoyo económico dado al estudiante Brayan Alexis Chavarro a través del programa JÓVENES INVESTIGADORES E INNOVADORES PARA GRUPOS DE INVESTIGACIÓN ITM 2020.

REFERENCIAS

- [1] R. C. E. Dias, P. Valderrama, P. H. Março, M. B. dos Santos Scholz, M. Edelmann y C. Yeretian, "Quantitative assessment of specific defects in roasted ground coffee via infrared-photoacoustic spectroscopy," *Food Chem.*, vol. 255, pp. 132–138, 2018, <https://doi.org/10.1016/j.foodchem.2018.02.076>.
- [2] C. Lao, M. Iglesias y R. Rodríguez, *Introducción a las técnicas analíticas aplicadas a la determinación cualitativa y cuantitativa de compuestos y elementos químicos presentes en el medio ambiente*, 2006.
- [3] D. A. Skoog, D. M. West, J. Holler y S. Crouch, (PDF) *Libro Fundamentos Analitica Skoog 9ed* | Marisela Marquez Academia.edu. 2015.
- [4] D. M. A. Molina-Quijada, L. A. Medina-Juárez, G. A. González-Aguilar, R. M. Robles-Sánchez y N. Gámez-Meza, "Compuestos fenólicos y actividad antioxidante de cáscara de uva (*Vitis vinifera L.*) de mesa cultivada en el noroeste de México Phenolic compounds and antioxidant activity of table grape (*Vitis vinifera L.*) skin from northwest Mexico," *CyTA J. Food*, vol. 8, n. 1, pp. 57–63, May 2010. <https://doi.org/10.1080/19476330903146021>.
- [5] Ó. J. Medina Vargas, L. C. A. Garzón Salcedo y W. E. Espinosa Manrique, "Efecto del tiempo sobre la cuantificación de polifenoles totales con el reactivo de Folin-Ciocalteu," *I3+*, vol. 3, n. 2, p. 10, 2017. <https://doi.org/10.24267/23462329.216>.
- [6] Ó. A. Muñoz-Bernal et al., "Nuevo acercamiento a la interacción del reactivo de Folin-Ciocalteu con azúcares durante la cuantificación de polifenoles totales," *TIP Revista Especializada en Ciencias Químico-Biológicas*, vol. 20, n. 2, pp. 23–28, Jul. 2017. <https://doi.org/10.1016/j.recqb.2017.04.003>.
- [7] F. D. S. Grasel, M. F. Ferrão y C. R. Wolf, "Development of methodology for identification the nature of the polyphenolic extracts by FTIR associated with multivariate analysis," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 153, pp. 94–101, 2016. <https://doi.org/10.1016/j.saa.2015.08.020>.
- [8] Grupo de Quimiometría y Cualimetría de Tarragona, "Quimiometría, una disciplina útil para el análisis químico". 2020.
- [9] M. Salazar, P. López de Alba, A. Piñeyro López y N. Waksman de Torres, "Quimiometría aplicada al control de calidad de fármacos," *Cienc. UANL*, vol. 6, n. 3, 2012. <http://eprints.uanl.mx/id/eprint/1473>.
- [10] A. Rohman, Y. B. Che Man, P. Hashim y A. Ismail, "FTIR spectroscopy combined with chemometrics for analysis of lard adulteration in some vegetable oils," *CYTA J. Food*, vol. 9, n. 2, pp. 96–101, 2011. <https://doi.org/10.1080/19476331003774639>.
- [11] A. Stöbener, U. Naefken, J. Kleber y A. Liese, "Determination of trace amounts with ATR FTIR spectroscopy and chemometrics: 5-(hydroxymethyl)furfural in honey," *Talanta*, vol. 204, pp. 1–5, 2019. <https://doi.org/10.1016/j.talanta.2019.05.092>.
- [12] C. Christou, A. Agapiou y R. Kokkinofita, "Use of FTIR spectroscopy and chemometrics for the classification of carobs origin," *Journal of Advanced Research*, vol. 10, pp. 1–8, 2018. <https://doi.org/10.1016/j.jare.2017.12.001>.
- [13] R. C. E. Dias, P. Valderrama, P. H. Março, M. B. Dos Santos Scholz, M. Edelmann y C. Yeretian, "Data on roasted coffee with specific defects analyzed by infrared-photoacoustic spectroscopy and chemometrics," *Data in Brief*, vol. 20, pp. 242–249, Oct. 2018. <https://doi.org/10.1016/j.dib.2018.08.013>.

- [14] S. de la F. Fernandez, “Componentes Principales”. 2011.
- [15] J. A. Rodrigo, “Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE”. 2017. [Online]. Available: https://www.cienciadedatos.net/documentos/35_principal_component_analysis.
- [16] C. Gil Martinez, “RPubs Análisis de componentes principales (PCA)”. 2018 [Online]. Available: https://rpubs.com/Cristina_Gil/PCA.
- [17] L. Scrucca, M. Fop, T. B. Murphy y A. E. Raftery, “Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models,” *The R journal*, vol. 8, n. 1, pp. 289–317, 2016. <https://doi.org/10.32614/rj-2016-021>.