

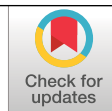


PREDICCIÓN DEL ORIGEN REGIONAL DEL CAFÉ DE COLOMBIA A PARTIR DE LA TÉCNICA DE ESPECTROSCOPIA DE INFRARROJO CERCANO - NIRS

Claudia Rocío Gómez Parra *, Aristóteles Ortiz **, Valentina Osorio Pérez ***

Gómez, C. R., Ortiz, A., & Osorio, V. (2022). Predicción del origen regional del café de Colombia a partir de la técnica de espectroscopia de infrarrojo cercano - NIRS. *Revista Cenicafé*, 73(2), e73205. <https://doi.org/10.38141/10778/73205>



Colombia es el tercer país productor de café y la primera fuente de cafés suaves lavados en el mundo, manteniéndose como un actor clave del mercado internacional. Aunque el café de Colombia ha logrado su consolidación por los atributos de calidad del grano, es necesario desarrollar herramientas que soporten de manera objetiva la identidad de los diversos orígenes al interior del país, debido a que las condiciones ambientales cambian a lo largo de las zonas geográficas, generando diferentes características de taza como resultado de las particularidades de su producción. Esta investigación evaluó la técnica NIRS para la predicción del origen de muestras de café almendra verde de ocho departamentos productores de Colombia. Para desarrollar los modelos de calibración, el método estadístico discriminante utilizado fue RMS (Root Mean Square) X Residualidad, utilizando la corrección Scatter Variación Normal estándar (SVN) y Detrend con el tratamiento matemático 2,4,4,1 (derivada, gap y smooth), para disminuir el ruido espectral generado por las características de la muestra. Los resultados evidenciaron un acierto medio global de clasificación del 92%, destacándose los departamentos de Antioquia y Nariño, y la zona de Sierra Nevada, con un 98%, 99% y 97%, respectivamente. El modelo desarrollado por zonas de producción (Zona Norte, Centro y Sur) presentó un acierto medio global de 89%. Los resultados de la investigación confirman que la técnica NIRS permite predecir el origen regional del café colombiano.

Palabras clave: Espectro, origen, regional, predicción, calibración, Cenicafé, Colombia.

REGIONAL ORIGIN PREDICTION OF COLOMBIAN COFFEE FROM THE NEAR-INFRARED SPECTROSCOPY TECHNIQUE - NIRS

Colombia is the third largest coffee-producing country and the first source of mild-washed coffees in the world, remaining a key player in the international market. Although Colombian coffee has achieved its consolidation due to the quality attributes of the bean, it is necessary to develop tools that objectively support the identity of the various origins within the country, because environmental conditions change throughout the geographic zones, generating different cup characteristics as a result of the particularities of its production. This research evaluated the NIRS technique for the prediction of the origin of green coffee samples from eight producing departments in Colombia. To develop the calibration models, the discriminant statistical method used was RMS (Root Mean Square) X Residuality, using the standard Normal Variation Scatter (SVN) and Detrend correction with the mathematical treatment 2,4,4,1 (derivative, gap, and smooth), to decrease the spectral noise generated by the sample characteristics. The results showed a global average classification accuracy of 92%, highlighting the departments of Antioquia and Nariño, and the Sierra Nevada area, with 98%, 99%, and 97%, respectively. The model developed by production areas (North, Central, and South Zone) presented a global mean accuracy of 89%. The research results confirm that the NIRS technique allows the prediction of the regional origin of Colombian coffee.

Keywords: Spectrum, origin, regional, prediction, calibration, Cenicafé, Colombia.

* Asistente de Investigación. Disciplina de Calidad. Centro Nacional de Investigaciones de Café, Cenicafé. <https://orcid.org/0000-0002-0685-8337>

** Investigador Científico I. Disciplina de Fisiología Vegetal, Cenicafé. <https://orcid.org/0000-0002-3242-1948>

*** Investigador Científico I. Disciplina de Calidad, Centro Nacional de Investigaciones de Café, Cenicafé. <https://orcid.org/0000-0002-1166-0165>



La técnica de espectroscopia de infrarrojo cercano (NIRS, por sus siglas en inglés), ha logrado grandes avances, posicionándose como una técnica precisa, confiable y reproducible para análisis cualitativos y cuantitativos en diferentes sectores (García-Olmo, 2009; Osborne et al., 1993; Williams & Norris, 1987). Algunas de las ventajas que ofrece es que permite obtener resultados rápidos, es una técnica no destructiva, que no requiere una transformación del material de análisis y permite analizar un gran número de muestras en un período corto de tiempo, presentando una alta reproducibilidad respecto al método de referencia, facilita el análisis de diferentes parámetros y realiza clasificaciones respecto a características cualitativas (Arndt et al., 2020; Cozzolino, 2002; García-Olmo, 2009; Pérez-Marín et al., 2007).

Esta técnica NIRS se fundamenta en la emisión de un haz de luz sobre la muestra, la cual dependiendo de su composición absorbe una cantidad determinada de radiación; la absorción depende principalmente de los enlaces presentes en sus moléculas (-CH, -NH, -OH, -CO). Este fundamento es conocido como la ley de Beer-Lambert, el cual establece que el valor de absorbancia de una muestra que contenga una sustancia con capacidad de absorber radiación NIR es directamente proporcional a la concentración de la sustancia (Alomar & Fuchslocher, 1998; Bertrand & Dufour, 2006; García-Olmo, 2009; Shenk & Westerhaus, 1991; Williams & Norris, 1987). Los valores de absorbancia a las diferentes longitudes de onda permiten obtener una curva denominada espectro NIR, que es el resultado de diferentes sobre tonos, bandas de combinación y absorciones electromagnéticas (Bertrand & Dufour, 2006; García-Olmo, 2009; Pérez-Marín et al., 2007; Shenk & Westerhaus, 1991).

Para analizar la información espectral y extraer la mayor cantidad de información posible

y extender el conocimiento del sistema químico se requiere de la quimiometría, la cual es explicada como la aplicación de las matemáticas en la química analítica (Mongay Fernández, 2005; Jiménez Torres, 2007; Vásquez et al., 2004). Las herramientas quimiométricas utilizadas con mayor frecuencia son los métodos de análisis multivariable, que se emplean para estimar cualquier propiedad física o química de la muestra a partir de múltiples variables espectrales, como los valores de absorbancia a diferentes longitudes de onda (Martens & Næs, 1992; Otto, 2017). Con base en lo anterior, en la tecnología NIRS se denomina etapa de calibración al proceso que se realiza en el desarrollo de la relación matemática entre los valores de absorbancia y los valores de las propiedades (químicas o físicas), al modelo matemático que se genera se le denomina ecuación de calibración y al grupo de muestras colectivo de calibración (García-Olmo, 2009; Williams, P.C., Sobering, D.C., 1996).

Los análisis que se utilizan para obtener la información requerida son cualitativo y cuantitativo. El cuantitativo es el análisis más utilizado, para lo cual se relacionan los valores de absorbancia con la concentración del analito, determinado previamente por un método de referencia, obteniendo la ecuación de calibración, la cual permite estimar su concentración (García-Olmo, 2009; Pérez-Marín et al., 2007). Con relación al análisis cualitativo, se define como la comparación de espectros de muestras desconocidas con grupos de espectros de características conocidas (Mark, 2007). Este tipo de análisis busca generar modelos de clasificación para un grupo o clases de determinadas muestras (Bertrand & Dufour, 2006; Deming et al., 1988; García-Olmo, 2009). Para el desarrollo de los modelos, se relaciona la información de datos espectrales y la información de la variable, la cual puede ser origen, variedad o calidad, entre otros; estableciéndose entre ellos una

función que permita correlacionar los dos tipos de información. Los métodos de clasificación más utilizados y de mayor aplicación a datos espectrales NIR son el análisis de componentes principales, la regresión de mínimos cuadrados parciales, el análisis factorial discriminante y discriminante X Residual.

Con relación al análisis discriminante, diferentes estudios evidencian resultados que confirman que a partir del espectro NIR es posible realizar la clasificación de diferentes grupos, a partir de variables o características categóricas. En la clasificación de predicción del origen se evidencian trabajos como el del loto blanco chino, en el cual utilizaron muestras de cuatro zonas productoras, desarrollando un modelo de calibración discriminante, con aciertos entre el 90% y el 97% (Xu et al., 2015). Para la clasificación por origen de *Junglan regia* L. (nueces), Arndt et al. (2020) tomaron muestras de siete países productores, desarrollaron un modelo de análisis discriminante lineal – LDA, con una precisión de clasificación del 77,1%. Para la clasificación de especies de *Panax* se tomaron muestras de tres zonas productoras, desarrollaron un modelo discriminante de mínimos cuadrados – PL, con un acierto de clasificación del 100% (Chen et al., 2011). En la identificación de diferentes marcas de licores, con el mismo sabor y el mismo contenido de alcohol, se desarrolló un modelo discriminante de mínimos cuadrados parciales-PLS con el cual obtuvieron una clasificación del 98% (Yang et al., 2017).

En Honduras, Giraudo et al. (2019) desarrollaron para café verde dos modelos de clasificación discriminante a partir del análisis de mínimos cuadrados parciales (PLS), los cuales clasificaban por continente y países, obteniendo un 98% y un 100% de clasificación, respectivamente. En Budapest, Alonso-Salces et al. (2009) desarrollaron el modelo de calibración a partir del PLS para

la clasificación de granos de café robusta de origen de Camerún, Vietnam e Indonesia, con un 94% de acierto. En Indonesia realizaron tres tipos de modelos de clasificación utilizando muestras de café robusta de tres regiones productoras, aplicaron el pretratamiento con variación normal estándar–SNV y *multiple scatter correction*-MSC, el cual presentó una clasificación del 93% de acierto para cada uno (Siregar et al., 2020). En Brasil, Bona et al. (2017) realizaron un modelo de clasificación geográfica de diferentes genotipos de café arábica aplicando *support vector machines* (SVN), con una clasificación de acierto de 100%. En Colombia se desarrollaron los primeros modelos para la predicción de origen de cinco localidades utilizando la distancia Mahalanobis, con una clasificación de 83% (Villegas et al., 2014).

Esta investigación desarrolló dos modelos de predicción, uno por el origen de diez departamentos cafeteros de Colombia y el otro por zonas de producción, determinando el porcentaje de acierto y verificando el uso de la técnica NIRS para la discriminación de los diferentes orígenes del café.

MATERIALES Y MÉTODOS

Muestras de café

Las muestras de café almendra verde utilizadas en la investigación provenían de los diez principales departamentos productores de café del país, en el caso de Sierra Nevada, esta zona se conformó por los departamentos de Cesar, La Guajira y Magdalena, configurando de esta manera ocho departamentos para el modelo (Tabla 1).

La selección de las muestras se realizó a partir de un muestreo aleatorio por conveniencia, con el objetivo de garantizar que se contara con muestras de café producidas en los límites

geográficos departamentales, al igual que al interior de cada departamento. Las muestras se tomaron directamente en las fincas cafeteras de cada uno de los departamentos, con el fin de asegurar la trazabilidad de las mismas. Estas muestras fueron procesadas de acuerdo a las condiciones establecidas por cada caficultor; una vez recibidas las muestras en el laboratorio se verificó el contenido de humedad, el cual debía estar en un rango del 10% al 12%; la humedad fue verificada con un medidor de humedad marca Kett PM420.

En total se analizaron 2.015 muestras, de las cuales se seleccionaron de forma aleatoria 1.762 para los modelos y se reservaron 253 para la validación cruzada, la distribución de las muestras se presenta en la Tabla 1, la cual contiene por cada uno de los departamentos el número total de muestras utilizadas para el modelo y la última columna las de validación.

Para realizar el análisis por la técnica de espectroscopia de infrarrojo cercano NIRS, las muestras de café pergamino seco (cps) fueron trilladas y se les realizó el análisis físico separando la almendra sana y la defectuosa. Posteriormente se pesaron 110 gramos de almendra sana para ser analizadas por NIRS.

Técnica Espectroscopia de Infrarrojo cercano – NIRS

Las muestras se analizaron en el equipo NIRS XDS RCA (2012) de FOSS, ubicado en el laboratorio de Cenicafé en Planalto (Manizales, Caldas). Se colocaron 110 g de cada una de las muestras de café almendra verde (entero) en una cubeta rectangular, se leyeron por duplicado, en un rango de longitud de onda de 400 nm – 2.490 nm, en un incremento de cada 0,5 nm para la captura del espectro; en total se realizaron

Tabla 1. Departamentos estudiados, número total de muestras analizadas y distribución de muestras para el modelo y para validación.

Departamento	Total de muestras	Muestras para desarrollo del modelo	Muestras para validación del modelo
Antioquia	148	110	38
Caldas	132	107	25
Cauca	261	226	35
Huila	265	235	30
Nariño	347	307	40
Santander	227	202	25
Sierra Nevada	454	414	40
Tolima	181	161	20
Total	2.015	1.762	253

4.030 lecturas. Posteriormente, con el software WinISI (versión 4 Foss Infracsoft International, USA) incorporado en el equipo, se analizó la información espectral y se desarrollaron los modelos de predicción discriminantes.

RESULTADOS Y DISCUSIÓN

Análisis espectral

A las muestras analizadas por la Técnica NIRS, se les realizó el análisis de absorbancia media por departamento, en la Figura 1 se observa que presentan la misma tendencia. Los departamentos de Caldas, Cauca y Nariño presentan menor valor de absorbancia respecto a Tolima, Santander y Sierra Nevada a partir de la longitud de 1.400.

Para el desarrollo de los modelos de calibración, es necesario realizar el análisis de componentes principales (PCA), el cual reduce la dimensionalidad de las variables, permite la identificación de muestras atípicas (*outlier*) y las discrimina en el caso que las muestras presenten este comportamiento (Chifarelli et al.,

2017; Cozzolino, D., et al, 2006); los parámetros estadísticos del análisis PCA estuvieron dados principalmente por el valor global de $H > 3,0$, el tratamiento matemático 3,3,3,1 y se analizó la información de la longitud de onda de 400 – 1.098,2 y 1.100 – 2.498,2, en total se identificaron 41 muestras atípicas, las cuales fueron retiradas, es decir, no se utilizaron para el desarrollo del modelo de calibración, ni para la validación.

En la Figura 2 se presenta la gráfica generada a partir del análisis PCA, donde el color rojo representa el grupo 1, conformado por Sierra Nevada y Santander, en color fucsia el grupo 2 conformado por Cauca y Huila y color verde el grupo 3 conformado por Caldas. Se observa que tienden a agruparse por zonas los departamentos evaluados; estas agrupaciones coinciden con las zonas de producción relacionadas con las épocas de cosecha (principal) que se encuentran referenciadas en el Boletín Agrometeorológico Cafetero No. 43, del mes de marzo de 2019 (Centro Nacional de Investigaciones de Café, 2019).

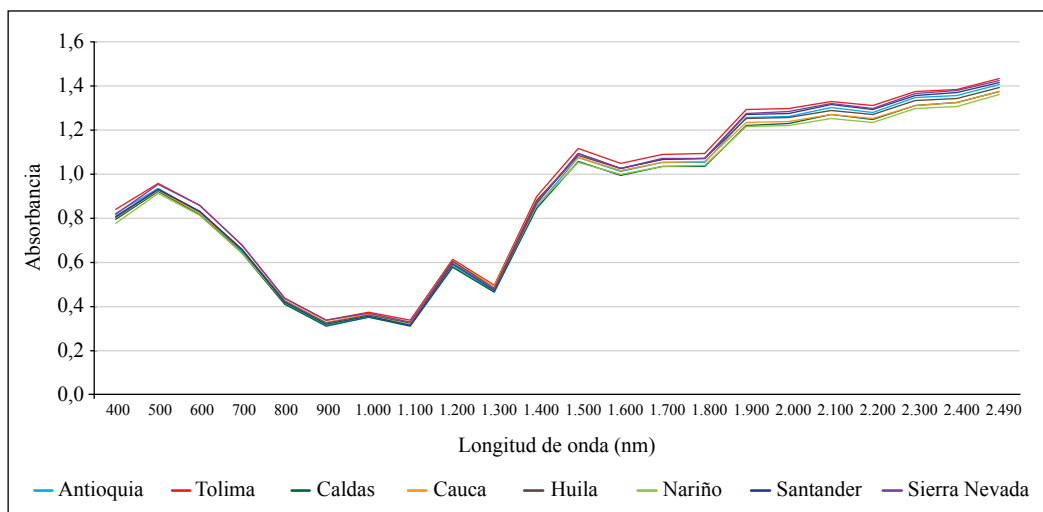


Figura 1. Absorbancia media del café verde, por departamento, respecto a la longitud de onda (400 a 2.490 nm).

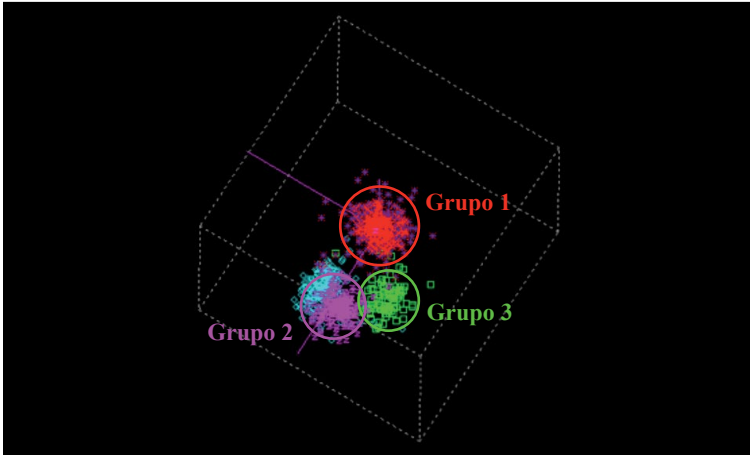


Figura 2. Gráfico de PCA para cuatro departamentos y la zona Sierra Nevada.

Con base en la tendencia de agrupación, se realizó el análisis por zonas, donde la zona Norte está conformada por Santander y Sierra Nevada (Figura 3); zona Centro (Figura 4) por Antioquia, Tolima y Caldas; y zona Sur (Figura 5) por Huila, Cauca y Nariño. En la Figura 6 se representan las tres zonas identificadas, con una clara separación entre estas ellas.

Desarrollo del modelo de predicción por departamento para café almendra

El modelo predictivo para este tipo de análisis son los denominados cualitativos, el cual se realiza a partir del análisis estadístico discriminante, y tiene como objetivo encontrar la combinación lineal de las variables independientes que permitan la máxima separación de las clases (Bolaños Alfaro, 2016; Prades et al., 2013; Ruiz-Aranibar, 2013). Teniendo en cuenta el análisis espectral presentado y evidenciando el comportamiento espectral de las muestras se desarrollaron los siguientes modelos de predicción:

1. Por departamentos. Se realizó un modelo para predecir individualmente los ocho departamentos evaluados para café almendra.

2. Por zonas. Se agruparon los departamentos en tres zonas (Norte, Centro y Sur) y el modelo debía predecir a qué zona pertenecía para café almendra.

Para los modelos de predicción desarrollados se realizaron diferentes pruebas con las herramientas estadísticas discriminantes ofrecidas por el Software WinISI, como mínimos cuadrados parciales (PLS), por correlación, máxima distancia, distancia mahalanobis, máxima residualidad y por RMS X residualidad, siendo este último el que presentó un mejor resultado, es decir, mayor porcentaje de acierto, lo que lo convierte en el criterio principal para la selección de los parámetros estadísticos; en la Tabla 2 se presenta el porcentaje de acierto de clasificación por cada una de los métodos estudiados.

Para dar claridad a la interpretación de los resultados, cuando se desarrolla el modelo de calibración y la posterior validación, el software genera la matriz de confusión, una herramienta estadística que permite hacer el análisis de observaciones emparejadas (Ariza-López et al., 2018; Sánchez-Muñoz, 2016), como se observa en la Tabla 3

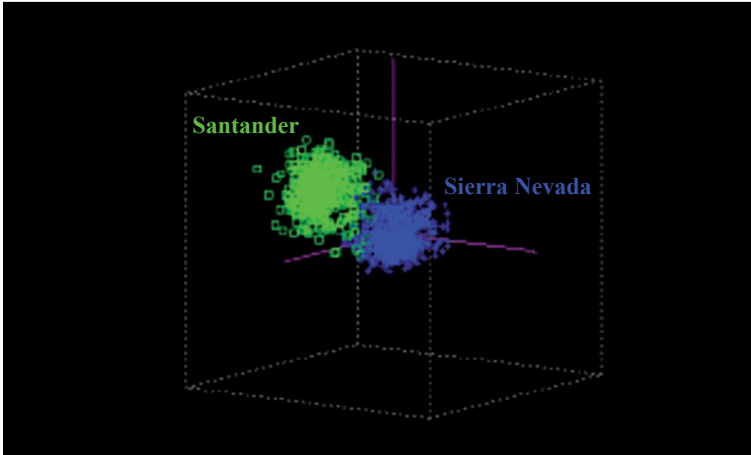


Figura 3. Gráfico de PCA Zona Norte Santander y Sierra Nevada.

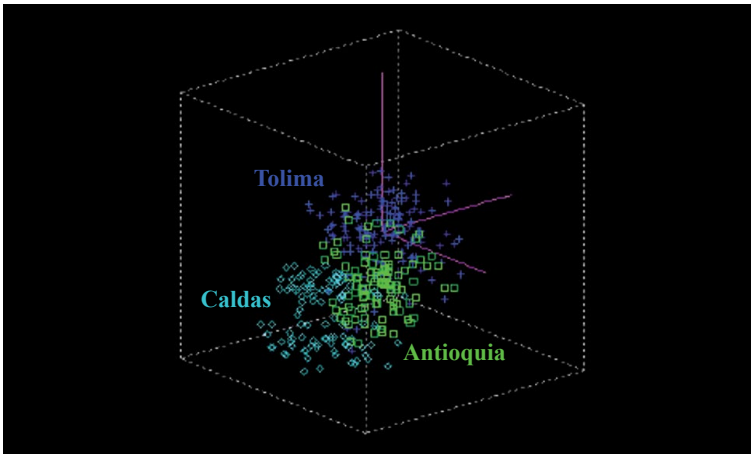


Figura 4. Gráfico de PCA Zona Centro, Tolima, Antioquia y Caldas.

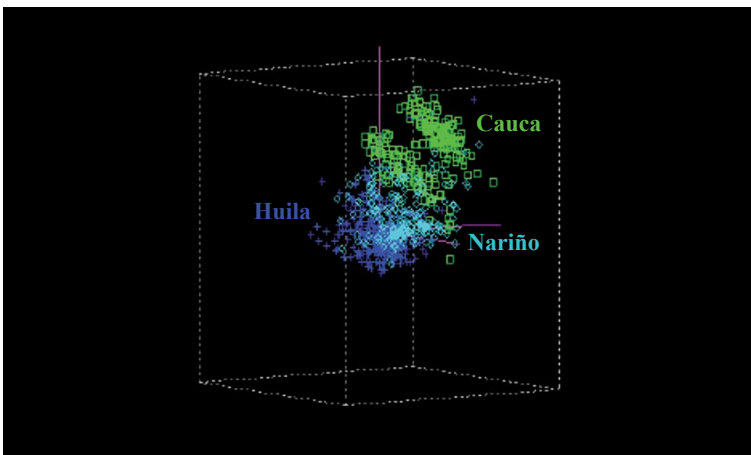


Figura 5. Gráfico de PCA Zona Sur, Huila, Cauca y Nariño.

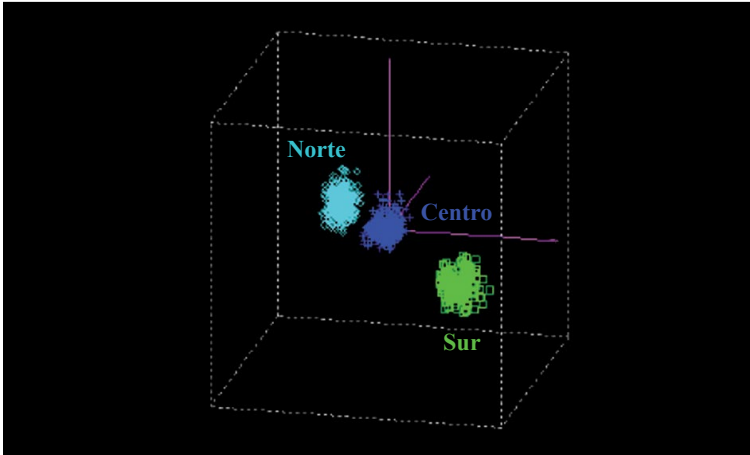


Figura 6. Gráfico de PCA Zona Sur, Centro y Norte.

Tabla 2. Porcentaje de acierto por cada uno de los métodos estadísticos discriminantes por departamento o zona, para predecir el origen del café almendra.

Departamento	PLS	Correlación	Máxima Distancia	Distancia Mahalanobis	Máximo X Residual	RMS X Residual
Antioquia	0,0	50,0	36,4	20,0	52,7	98,0
Caldas	0,0	13,1	37,4	51,4	56,1	83,0
Cauca	17,1	45,4	59,0	31,2	50,5	93,0
Huila	0,0	8,5	21,3	6,8	31,5	88,0
Nariño	2,9	62,5	16,0	42,7	51,1	99,0
Santander	0,0	17,8	6,4	3,0	42,1	87,0
Sierra Nevada	30,0	53,9	41,5	45,9	56,5	97,0
Tolima	0,0	38,5	56,5	45,3	39,8	91,0

los verdaderos positivos (color verde) corresponden a las muestras que clasifican correctamente, es decir, al departamento del cual tienen su origen, y los falsos positivos (color rojo) corresponden a las muestras que clasificaron de forma incorrecta, permitiendo establecer el porcentaje global de acierto

a través de indicadores de sensibilidad por departamento. Este mismo procedimiento se realizó para la validación del modelo con el set de muestras reservadas para ese objetivo, generando la matriz de confusión de confirmación de desempeño del modelo desarrollado (Tabla 4).

Modelo de clasificación por departamento

El modelo que presentó el mejor resultado fue el desarrollado a partir de método discriminante RMS X Residual; se ingresaron ocho grupos de variables, cada una correspondiente a un departamento o zona (Antioquia, Caldas, Cauca, Huila Nariño, Santander, Sierra Nevada y Tolima), la variación explicada del modelo fue de 0,99, tolerancia de 100, la corrección de la dispersión se realizó con *standard normal variate*-SNV y Drend -DT, donde el SNV redujo el efecto del tamaño de la partícula y DT corrigió la curvatura lineal, el tratamiento matemático fue 2,4,4,1 (derivada, segmento de onda, suavizado 1, suavizado 2). En la matriz de confusión (Tabla 3) de café almendra verde, el modelo de calibración discriminante presentó:

- Un porcentaje de acierto global de 92% y un error global de 8%.
- Los departamentos de Nariño y Antioquia, y la zona Sierra Nevada presentaron los mayores porcentajes de acierto, 99%, 98% y 97%, con respecto a los demás departamentos.
- Los departamentos que presentaron un mayor número de falsos positivos fueron Caldas, Santander y Huila, con 17%, 13% y 12%, respectivamente.
- El mayor número de falsos positivos de los departamentos de Cauca, Huila y Nariño, están ubicados en estos mismos departamentos. En el caso de los departamentos de Cauca y Huila el 86% y el 57% de estos falsos positivos se encuentran ubicados en Nariño, lo que mostraría consistencia con lo observado en el PCA;

es decir, que la huella espectral presenta un comportamiento muy similar para estos tres departamentos.

Matriz de validación cruzada por departamento

La validación cruzada se realizó con el grupo de muestras (253) reservadas para la validación del modelo. En la Tabla 4 se presenta la matriz de confusión obtenida después de aplicar el modelo desarrollado por departamentos donde Caldas y Tolima reducen su porcentaje de acierto en un 3% y un 11%, respectivamente, aunque puede observarse la misma tendencia de clasificación del modelo.

Desarrollo del modelo de predicción por zonas

Las zonas fueron conformadas por los departamentos seleccionados de acuerdo a los resultados obtenidos en los grupos (Figura 6).

- Zonas Norte: Santander y Sierra Nevada (Cesar, La Guajira y Magdalena).
- Zona Centro: Antioquia, Caldas y Tolima.
- Zona Sur: Huila, Cauca y Nariño.

Los parámetros matemáticos que presentaron un mejor resultado de clasificación del modelo discriminante desarrollado se describen en la Tabla 5.

Al realizar el modelo por zonas puede evidenciarse en la matriz de confusión en la Tabla 6, que el modelo presentó una clasificación media global de 89,2% y un error medio global de 10,8%, la zona Sur presentó una mejor clasificación respecto a las otras zonas evaluadas.

Tabla 3. Matriz de confusión por departamento y zona del modelo predictivo.

Depto.	Antioquia	Caldas	Cauca	Huila	Nariño	Santander	Sierra Nevada	Tolima	Acierto	Error	Total n	Acierto (%)
Antioquia	108	0	0	0	1	0	0	1	108	2	110	98
Caldas	2	89	0	0	10	6	0	0	89	18	107	83
Cauca	0	0	211	1	13	1	0	0	211	15	226	93
Huila	0	0	10	207	16	1	0	1	207	28	235	88
Nariño	0	0	1	0	304	1	0	1	304	3	307	99
Santander	1	0	8	1	16	176	0	0	176	26	202	87
Sierra Nevada	2	0	1	1	2	4	403	1	403	11	414	97
Tolima	2	0	1	4	8	0	0	146	146	15	161	91

Verdaderos positivos (color verde); falsos positivos (color rojo)

Tabla 4. Matriz de la validación cruzada del modelo desarrollado por departamento y zona.

Depto.	Antioquia	Caldas	Cauca	Huila	Nariño	Santander	Sierra Nevada	Tolima	Total n	Acierto (%)
Antioquia	38	0	0	0	0	0	0	0	38	100
Caldas	1	20	0	0	3	1	0	0	25	80
Cauca	0	0	34	0	1	0	0	0	35	97
Huila	0	0	0	26	4	0	0	0	30	87
Nariño	0	0	0	0	40	0	0	0	40	100
Santander	0	0	1	1	1	22	0	0	25	88
Sierra Nevada	0	0	1	0	0	0	39	0	40	98
Tolima	1	0	1	1	1	0	0	16	20	80

Tabla 5. Características matemáticas del modelo predictivo por zonas.

Descripción	Zona Sur	Zona Centro	Zona Norte
Método discriminante	RMSXResiduals	RMSXResiduals	RMSXResiduals
Número de factores	3	3	2
Variación explicada del modelo	0,99	0,99	0,99
Tolerancia	100	100	100
Corrección Scatter	SNV_DETREND	SNV_DETREND	SNV_DETREND
Tratamiento matemático	2,4,4,1	2,4,4,1	2,4,4,1
Longitud de onda	408,1092,8	408,1092,8	408,1092,8
	1108,2492,8	1108,2492,8	1108,2492,8

Tabla 6. Matriz de clasificación del modelo desarrollado para las zonas Centro, Norte y Sur.

Zona	Centro	Norte	Sur	Acierto	Error	Total	Acierto (%)
Centro	373	30	49	373	79	452	82,5
Norte	9	478	46	478	55	533	89,7
Sur	15	21	732	372	36	768	95,3

Verdaderos positivos (color verde); falsos positivos (color rojo)

Matriz de validación cruzada por zonas

Al realizar la validación cruzada, la zona Centro presentó un menor porcentaje de acierto (65,1%) respecto a las otras zonas evaluadas, 81,5% para la zona Norte y 96,2% para la

zona Sur, las muestras que no clasificaron correctamente fueron principalmente las de origen Tolima (Tabla 7).

Los resultados obtenidos para los dos modelos de predicción del origen regional

Tabla 7. Matriz de validación cruzada del modelo desarrollado por zonas.

Zona	Centro	Norte	Sur	Total	Acierto (%)
Centro	54	12	17	83	65,1
Norte	1	53	11	65	81,5
Sur	1	3	101	105	96,2

y por zonas para café almendra, confirman la efectividad de la técnica NIRS para realizar la clasificación regional de los diez departamentos objeto de esta investigación. Trabajos referenciados por otros autores obtuvieron resultados similares como los realizados por Siregar et al. (2020), quienes obtuvieron un 93% de acierto en los modelos desarrollados para la clasificación de café robusta de tres regiones productoras de Indonesia; Giraudo et al. (2019) obtuvieron un 98% y 100% de acierto para los modelos de clasificación por países y continentes de café verde; Bona et al. (2017) realizaron un modelo para la clasificación geográfica en Brasil para el café arábica con un 80% de acierto, y en Budapest, Alonso – Salces et al. (2009) desarrollaron un modelo de clasificación de café robusta proveniente de Camerún, Vietnam e Indonesia, con un 98% de acierto.

Puede concluirse que los resultados confirman que la técnica de espectroscopia

de infrarrojo cercano NIRS es una técnica confiable y reproducible que permite la clasificación regional del café producido en los departamentos de estudio. El modelo predictivo de origen para ocho departamentos, presentó una clasificación media global de 92% y un error medio global de 8%. En la identificación del comportamiento espectral de las muestras evaluadas se evidenció una clara separación zonal del café de Colombia, razón por la que se desarrollaron los modelos por zonas Norte, Centro y Sur, donde el modelo presentó una clasificación media global de 89,2% y un error medio global de 10,8%.

AGRADECIMIENTOS

Al doctor Álvaro Gaitán Director de Cenicafé, a los Comités Departamentales Cafeteros, Líderes de Extensión, Servicio de Extensión, Almacafé, Cafecert y al personal del apoyo Anderson Arias y Víctor Castañeda. Fuente de financiación: recursos propios, CAL104012.

LITERATURA CITADA

Alomar, D., & Fuchslocher, R. (1998). Fundamentos de la espectroscopia de reflectancia en el infrarrojo cercano (NIRS) como método de análisis de forrajes. *Agro sur*, 26(1), 88–104. <https://doi.org/10.4206/agrosur.1998.v26n1-11>

Alonso-Salces, R. M., Serra, F., Reniero, F., & Héberger, Ká. (2009). Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): Chemometric evaluation of phenolic and methylxanthine contents. *Journal of Agricultural*

- and Food Chemistry, 57(10), 4224–4235. <https://doi.org/10.1021/jf8037117>
- Ariza-López, F. J., Rodríguez-Avi, J., & Alba-Fernández, V. (2018). Control estricto de matrices de confusión por medio de distribuciones multinomiales. *GeoFocus. International Review of Geographical Information Science and Technology*, 21, 215–226. <https://doi.org/10.21138/GF.591>
- Arndt, M., Drees, A., Ahlers, C., & Fischer, M. (2020). Determination of the geographical origin of walnuts (*Juglans regia* L.) using near-infrared spectroscopy and chemometrics. *Foods*, 9(12), 1860. <https://doi.org/10.3390/foods9121860>
- Bertrand, D., & Dufour, E. (Eds.). (2005). *La spectroscopie infrarouge et ses applications analytiques* (2a ed.). Tec et Doc–Lavoisier.
- Bolaños Alfaro, J. D. (2016). El método NIR combinado con el análisis quimiométrico PLS-da para determinar la adulteración del aceite de oliva con aceite de girasol. *Pensamiento Actual*, 16(26), 163. <https://doi.org/10.15517/pa.v16i26.25764>
- Bona, E., Marquetti, I., Link, J. V., Makimori, G. Y. F., da Costa Arca, V., Guimarães Lemes, A. L., Ferreira, J. M. G., dos Santos Scholz, M. B., Valderrama, P., & Poppi, R. J. (2017). Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. *LWT–Food Science and Technology*, 76, 330–336. <https://doi.org/10.1016/j.lwt.2016.04.048>
- Centro Nacional de Investigaciones de Café. (2019). *Boletín agrometeorológico cafetero* (Núm. 43). <https://agroclima.cenicafe.org/web/guest/boletin-agrometeorologico>
- Chen, X., Wu, D., He, Y., & Liu, S. (2011). Nondestructive differentiation of panax species using visible and shortwave near-infrared spectroscopy. *Food and Bioprocess Technology*, 4(5), 753–761. <https://doi.org/10.1007/s11947-009-0199-6>
- Chifarelli, V., Giménez, A. M., Nisgoski, S., & Moglia, J. G. (2017). Aplicación de la técnica NIRS (Espectroscopía de infrarrojo cercano) en la discriminación de la madera de Bulnesia sarmientoi. En A. M. Gimenez & J. G. Moglia (Eds.), *Los bosques actuales del chaco semiárido argentino ecoanatomía y biodiversidad: Una mirada propositiva*. (pp. 267–279). Universidad Nacional de Santiago del Estero. https://fcf.unse.edu.ar/archivos/publicaciones/libro-ecoanatomia/libro_Los-bosques-actuales-del-Chaco-semiarido-argentino-Ecoanatomia-y-biodiversidad-una-mirada-propositiva.pdf
- Cozzolino, D. (2002). Uso de la espectroscopía de reflectancia en el infrarrojo cercano (NIRS) en el análisis de alimentos para animales. *Agrociencia*, 6(2), 25–32. <http://www.fagro.edu.uy/~agrociencia/VOL6/2/p25-32.pdf>
- Cozzolino, D., Fernández, E. G., Restaino, E. A., & La Manna, A. (2006). Determinación de la composición química de heno mediante la espectrofotometría en el infrarrojo cercano (NIRS). *Revista Argentina de producción animal*, 26(1), 203–209. <http://www.ainfo.inia.uy/digital/bitstream/item/12191/1/4261-22466-1-PB.pdf>
- Deming, S. N., Michotte, Y., Massart, D. L., Kaufman, L., & Vandeginste, B. G. M. (1988). *Chemometrics: A textbook*. Elsevier Science.
- García-Olmo, J. (2009). *Clasificación y autenticación de canales de cerdo ibérico mediante espectroscopía en el infrarrojo cercano (NIRS)* [Tesis de Doctorado, Universidad de Córdoba]. <http://hdl.handle.net/10396/2340>
- Giraud, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control*, 99, 137–145. <https://doi.org/10.1016/j.foodcont.2018.12.033>
- Mark, H. (2007). Data analysis: multilinear regression and principal component analysis. En D. A. Burns & E. W. Ciurczak (Eds.), *Handbook of Near-Infrared Analysis* (3a ed., pp. 151–188). CRC Press. <https://doi.org/10.1201/9781420007374>
- Martens, H., & Næs, T. (1992). *Multivariate Calibration*. John Wiley & Sons.
- Mongay Fernández, C. (2005). *Quimiometría*. Universitat de València.
- Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis* (2nd ed). Longman Scientific & Technical; Wiley.
- Otto, M. (2017). *Chemometrics: Statistics and computer application in 4 analytical chemistry* (3a ed.). Wiley-VCH Verlag.
- Pérez-Marín, D., Garrido-Varo, A., & Guerrero, J. E. (2007). Non-linear regression methods in NIRS quantitative analysis. *Talanta*, 72(1), 28–42. <https://doi.org/10.1016/j.talanta.2006.10.036>
- Prades, C., García-Olmo, J., Romero Prieto, T., García de Ceca, J. L., & López-Luque, R. (2013). Aplicación de

- la tecnología de espectroscopia de infrarrojo cercano (NIRS) a la clasificación por calidad del corcho en plancha. *VI Congreso Forestal Español*, 2–13. http://secforestales.org/publicaciones/index.php/congresos_forestales/article/view/14270
- Ruiz-Aranibar, G. (2013). Análisis discriminante. *Revista Varianza*, 10(2), 1–10. <http://www.revistasbolivianas.ciencia.bo/pdf/rv/n10/n10a02.pdf>
- Sánchez-Muñoz, J. M. (2016). Análisis de Calidad Cartográfica mediante el estudio de la Matriz de Confusión. *Pensamiento matemático*, 6(2), 9–26. <https://dialnet.unirioja.es/servlet/articulo?codigo=5998855>
- Shenk, J. S., & Westerhaus, M. O. (1991). Populations Structuring of Near Infrared Spectra and Modified Partial Least Squares Regression. *Crop Science*, 31(6), 1548-1555. <https://doi.org/10.2135/cropsci1991.0011183X003100060034x>
- Siregar, S. D., Rindang, A., & Ayu, P. C. (2020). Principle Component Analysis (PCA)—Classification of Arabica green bean coffee of North Sumatera Using FT–NIRS. *IOP Conference Series: Earth and Environmental Science*, 454(1), 012046. <https://doi.org/10.1088/1755-1315/454/1/012046>
- Jiménez Torres, P.A. (2007). *Identificación de harinas de yuca (Manihot esculenta Crantz) con alto contenido proteico mediante espectroscopia de infrarrojo cercano (NIRS)* [Tesis de Maestría, Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/handle/unal/2473>
- Vásquez, D. R., Abadía, B., & Arreaza, L. C. (2004). Aplicación de la Espectroscopia de Reflectancia en el Infrarrojo Cercano (NIRS) para la caracterización nutricional del pasto Guinea y del grano de maíz. *Ciencia & Tecnología Agropecuaria*, 5(1), 49–55. https://doi.org/10.21930/rcta.vol5_num1_art:24
- Villegas, A. M., Pérez-Henao, C., Arana, V. A., Sandoval, T., Posada-Suárez, H. E., Garrido, A., Guerrero, J., Pérez-Marín, D., & García-Olmo, J. (2014). Identificación de origen y calibración para tres compuestos químicos en café, por espectroscopia de infrarrojo cercano. *Revista Cenicafé*, 65(1), 7–16. <https://biblioteca.cenicafe.org/handle/10778/551>
- Williams, P., & Norris, K. H. (1987). *Near-infrared technology in the agricultural and food industries*. American Association of Cereal Chemists. <https://doi.org/10.1002/food.19880320825>
- Williams, P. C., & Sobering, D. C. (1996). How Do We Do It: A Brief Summary of the Methods We Use in Developing near Infrared Calibration. In A. M. C. Davis & P. Williams (Eds.), *Near Infrared Spectroscopy: The Future Waves* (pp. 185–188). NIR Publications.
- Xu, L., Cai, C.-B., She, Y.-B., & Chen, L.-J. (2015). Discriminating the Geographical Origins of Chinese White Lotus Seeds by Near-Infrared Spectroscopy and Chemometrics. *Journal of Spectroscopy*, 2015, e831246. <https://doi.org/10.1155/2015/831246>
- Yang, B., Yao, L., & Pan, T. (2017). Near-Infrared Spectroscopy Combined with Partial Least Squares Discriminant Analysis Applied to Identification of Liquor Brands. *Engineering*, 9(2), 181–189. <https://doi.org/10.4236/eng.2017.92009>