

**Predicción del contenido de grasa en cacao a partir de espectroscopia infrarroja (FTIR-ATR) y métodos quimiométricos**

**Jennifer Steffens Ariza Ocaño**

**Karen Johanna Osorio Niño**

**Trabajo de grado como requisito parcial para optar al título de Química**

**Director del proyecto**

**Enrique Mejía Ospino**

**Codirector del proyecto**

**Luis Javier López Giraldo**

**Universidad Industrial de Santander**

**Facultad de Ciencias**

**Escuela de Química**

**Bucaramanga**

**2018**

### **Agradecimientos**

Expresamos nuestro profundo agradecimiento al doctor Enrique Mejía Ospino y al doctor Luis Javier López Giraldo por guiarnos y brindarnos su apoyo en la dirección de este proyecto.

Al profesor Jader Guerrero por su tiempo, colaboración y aporte de sus valiosos conocimientos.

Al Laboratorio de Espectroscopia Atómica y Molecular (LEAM) así como al Centro de Investigación en Ciencia y Tecnología de Alimentos (CICTA) por brindarnos los recursos necesarios para llevar a cabo este trabajo.

Agradecemos a nuestros padres y amigos que nos brindaron su apoyo y compañía en cada paso dado de nuestra carrera universitaria.

**Contenido**

	<b>Pág.</b>
Introducción .....	14
1. Antecedentes .....	16
2. Aspectos preliminares .....	18
2.1 Objetivos .....	18
2.1.1 Objetivo general .....	18
2.1.2 Objetivos específicos .....	18
3. Marco Teórico .....	19
3.1 El cacao en Colombia .....	19
3.1.1 Grasa en cacao .....	20
3.2 Espectroscopia infrarroja .....	21
3.2.1 Espectroscopia infrarroja con transformada de Fourier (FTIR) .....	21
3.3 Pretratamientos espectrales .....	23
3.3.1 Suavizado .....	23
3.3.2 Normalización .....	23
3.3.3 Derivación .....	24
3.4 Quimiometría .....	24
3.4.1 Métodos quimiométricos .....	25
3.4.1.1 Análisis de componentes principales (PCA) .....	26

3.4.1.2 Método de Mínimos Cuadrados Parciales .....	27
3.4.2 Calibración multivariable.....	29
3.4.1.3 Construcción de modelos de calibración multivariable .....	29
4. Metodología .....	30
4.1 Materiales y reactivos .....	31
4.1.1 Reactivos.....	31
4.1.2 Equipos .....	31
4.2 Selección de muestras .....	32
4.3 Preparación de muestras para la toma de espectros .....	34
4.4 Optimización de las condiciones experimentales .....	34
4.5 Pretratamiento de los espectros.....	35
4.5.1 Suavizado.....	35
4.5.2 Normalización.....	35
4.5.3 Primera derivada. ....	35
4.6 Análisis por componentes principales. ....	36
4.7 Construcción del modelo .....	36
4.8 Verificación de la viabilidad del modelo .....	37
5. Resultados .....	37
5.1 Parámetros instrumentales .....	37
5.1.1 Variación del número de SCAN. ....	37
5.1.2 Variación de la resolución espectral. ....	39
5.1.3 Variación de la velocidad del espejo .....	40
5.1.4 Verificación del rendimiento del equipo.....	41

---

5.2 Pretratamiento espectral.....	43
5.3 Análisis por componentes principales (PCA).....	46
5.4 Método de mínimos cuadrados parciales (PLS) .....	49
5.4.1 Determinación del número de componentes o variables latentes .....	49
5.4.2 Modelos propuestos .....	52
6. Conclusiones .....	59
7. Divulgación Científica .....	60
Referencias Bibliográficas .....	61

**Lista de Figuras**

	<b>Pág.</b>
<i>Figura 1.</i> Esquema de la metodología llevada a cabo. ....	31
<i>Figura 2.</i> Espectrómetro Infrarrojo utilizado en el desarrollo del proyecto. ....	32
<i>Figura 3.</i> Espectros de la muestra 63, tomados con diferente número de SCAN. ....	38
<i>Figura 4.</i> Espectros de la muestra 63, tomados con diferente resolución espectral. ....	39
<i>Figura 5.</i> Espectros de la muestra 63 tomados con diferentes valores de velocidad del espejo. .	40
<i>Figura 6.</i> Espectros de la muestra 56 obtenidos en la prueba de repetibilidad. ....	42
<i>Figura 7.</i> Espectros de las 33 muestras sin pretratamiento alguno.....	43
<i>Figura 8.</i> Espectros de las 33 muestras con el método de suavizado Gaussian Filter.....	44
<i>Figura 9.</i> Espectros de las 33 muestras normalizados.....	45
<i>Figura 10.</i> Espectros de las 33 muestras tratados con normalización SVN. ....	45
<i>Figura 11.</i> Derivada Savitzky-Golay aplicada a los espectros de las 33 muestras.....	46
<i>Figura 12.</i> Agrupación según el origen de las 33 muestras de cacao en la gráfica de Scores para el PC1 y PC2. ....	48
<i>Figura 13.</i> Agrupación según el porcentaje de grasa de las 33 muestras de cacao en la gráfica de Scores para el PC1 y PC2.....	48
<i>Figura 14.</i> T <sup>2</sup> de Hotelling aplicado a la gráfica de Scores de las 33 muestras de cacao. ....	49
<i>Figura 15.</i> Curva de tendencia de RMSEP contra RMSECal de predicción del porcentaje de grasa en cacao con el pretratamiento de: a) Normalización por área, b) Normalización z-score. 51	

*Figura 16.* (RMSEP) contra el número de componentes para los modelos con pretratamiento de: a) Normalización por área, b) Normalización z-score. .... 52

*Figura 17.* Curva de calibración para el modelo de predicción de grasa de cacao bajo un pretratamiento de: a) Normalización por área, b) Normalización z-score..... 55

*Figura 18.* Gráfica de los residuales frente a los valores reales de grasa de cacao para los modelos bajo: a) Normalización por área, b) Normalización z-scores ..... 57

*Figura 19.* Histogramas de los residuales para los modelos bajo: a) Normalización por área, b) Normalización z-scores ..... 58

**Lista de Tablas**

	<b>Pág.</b>
Tabla 1. Características estándar recomendadas por la A.O.C.S (American Oil Chemists' Society) para la manteca de cacao.....	21
Tabla 2. Porcentaje de grasa de treinta y tres muestras de cacao colombianos. ....	33
Tabla 3. Relación Señal/Ruido obtenidas con los espectros tomados a diferentes números de SCAN.....	38
Tabla 4. Relación Señal/Ruido obtenidas con los espectros tomados a diferente resolución espectral. ....	40
Tabla 5. Relación Señal/Ruido obtenidas con los espectros tomados a diferente velocidad del espejo. ....	41
Tabla 6. Parámetros establecidos para la toma de espectros de cada una de las muestras. ....	41
Tabla 7. Desviación estándar del % de transmitancia a 2915,68 cm <sup>-1</sup> y 1743 cm <sup>-1</sup> de las 20 repeticiones. ....	42
Tabla 8. Valores medidos y predichos para el set de muestras de validación en el modelo desarrollado a partir de normalización por área. Error cuadrático medio de predicción de 0,36 .....	56
Tabla 9. Valores medidos y predichos para el set de muestras de validación en el modelo desarrollado a partir de normalización SVN. Error cuadrático medio de predicción de 0,58 .....	56

## Resumen

**Título:** Predicción del contenido de grasa en cacao a partir de espectroscopia infrarroja (FTIR-ATR) y métodos quimiométricos\*

**Autores:** Ariza Ocaño Jennifer  
Osorio Niño Karen\*\*

**Palabras Claves:** Espectroscopia infrarroja (FTIR-ATR), Quimiometría, Grasa en cacao.

### Descripción

El contenido de grasa en granos de cacao es una de las características que los compradores de esta materia prima consideran a la hora de darle un valor al grano. De manera convencional, la determinación involucra varios pasos que involucran extracción, concentración del solvente y finalmente determinación gravimétrica. Estos pasos consumen tiempo y recursos representados en horas hombre y reactivos químicos. Otra opción a los métodos convencionales es el uso de técnicas quimiométricas que son capaces de predecir una propiedad con información que puede adquirirse en línea, minimizando así tiempo y esfuerzo sin sacrificar exactitud. En este trabajo se desarrollaron dos modelos predictivos quimiométricos para determinar el porcentaje de grasa total en el cacao. Para esto se utilizó la información espectral obtenida por espectroscopia infrarroja con transformada de Fourier acoplada a Reflectancia Total Atenuada (FTIR-ATR), en combinación con mínimos cuadrados parciales (PLS, por su nombre en inglés). Los modelos mostraron coeficientes de correlación mayor a 0,95 y errores de predicción de 0,36 y 0,58. Los resultados obtenidos son bastante prometedores, desde el punto de vista estadístico, para utilizarlos como alternativas de los métodos convencionales para la determinación de grasa en cacao. La aplicabilidad del método se evaluó utilizando parámetros estadísticos como son el error cuadrático medio de calibración y validación, RMSEC y RMSEV, además del coeficiente de regresión lineal de los modelos. Por los resultados se considera que esta es una metodología que requiere un menor tiempo y costo de análisis que los métodos convencionales, por los que puede ser aplicada para la rápida determinación de las propiedades físicoquímicas de los cacaos colombianos.

---

\* Trabajo de grado

\*\* Facultad de Ciencias. Escuela de Química. Director: Enrique Mejía Ospino. Codirector: Luis Javier López Giraldo.

## Abstract

**Title:** Prediction of fat content in cocoa from infrared spectroscopy (FTIR-ATR) and chemometric methods\*.

**Authors:** Ariza Ocaño Jennifer  
Osorio Niño Karen\*\*

**Key Words:** Infrared spectroscopy (FTIR-ATR), Chemometrics, Cocoa fat.

### Description

The fat content in cocoa beans is one of the characteristics that buyers of this raw material consider when it comes to giving a value to the grain. Conventionally, the determination involves several steps involving extraction, concentration of the solvent and finally gravimetric determination. These steps consume time and resources represented in man hours and chemical reagents. Another option for conventional methods is the use of chemometric techniques that are capable of predicting a property with information that can be acquired online, minimizing time and effort without sacrificing accuracy. In this work, it was developed in the chemometric predictive models to determine the percentage of total fat in cocoa. For this purpose, the information obtained was obtained by infrared spectroscopy with Fourier transform coupled to Attenuated Total Reflectance (FTIR-ATR), in combination with partial least squares (PLS, by its English name). The models experienced correlation coefficients greater than 0.95 and prediction errors of 0.36 and 0.58. The results were quite promising, from the statistical point of view, to use them as alternatives to conventional methods for the determination of fat in cocoa. The application of the method was evaluated using statistical parameters such as the mean squared error of calibration and validation, RMSEC and RMSEV, in addition to the coefficient of linear regression of the models. The results are considered to be a methodology that requires a shorter time and cost of analysis than conventional methods, for which it can be applied for the rapid determination of the physical properties of Colombian cocoas.

---

\* Grade work

\*\* Science Faculty School of Chemistry. Director: Enrique Mejia Ospino. Codirector: Luis Javier López Giraldo.

## Introducción

El cacao es uno de los productos agrícolas de mayor importancia en el país y particularmente en el departamento de Santander. En el 2017 Colombia logró un récord con una producción de 60.535 toneladas de cacao, con un aumento del 6,6% con relación al año anterior; las exportaciones a diferentes países también aumentaron en un 11%, según las cifras presentadas por Fedecacao se vendieron al exterior 11.688 toneladas (Fedecacao, 2017).

La grasa es uno de los componentes más importantes del cacao, puede corresponder hasta un 58% de los granos secos, está compuesta por triglicéridos y ácidos grasos como el esteárico, palmítico, oleico entre otros. Son estos ácidos grasos los que generan las propiedades fisicoquímicas únicas de la manteca de cacao (Naik & Kumar, 2014). El contenido de grasa del cacao tiene un valor importante industrialmente debido a que es extraída generalmente como manteca de cacao, una grasa saludable, utilizada como materia prima a nivel industrial como, por ejemplo, en la industria farmacéutica y en la industria de cosméticos, para la producción de diversos productos de belleza con propiedades hidratantes (Kalustian, 1985).

La grasa de cacao también es el componente principal en la ingesta de chocolate y en otros productos de confitería, debido a esto es una de las grasas vegetales más importantes y apreciadas en comparación con otras, esta es la razón para que su proceso de extracción y cuantificación sea uno de los más indispensables al momento de trabajar con cacao. Existen diversos métodos para extraer la grasa: por prensado utilizando prensas hidráulicas o mecánicas, prensas de tornillo y método de extracción por solvente. Por lo general la cuantificación se hace por procesos

convencionales de extracción usando solventes. El análisis se realiza utilizando el método de extracción Soxhlet siguiendo la norma “AOAC Official Method 963.15, Fat in Cacao Products”, esta técnica requiere el uso de grandes volúmenes de disolvente y largos tiempos de extracción, lo que encarece la prueba. Además, este procedimiento genera diversos riesgos para la salud debido a los diferentes solventes orgánicos que se utilizan. La eliminación de estos solventes también es un tema preocupante puesto que generan un efecto tóxico sobre el medio ambiente. Por esta razón se quiere buscar una alternativa para obtener un método más rápido, limpio y amigable con el medio ambiente. (Di Khanh, 2015)

La espectroscopia infrarroja por transformada de Fourier (FTIR) es una técnica de análisis rápido que mide las vibraciones de los enlaces dentro de grupos funcionales, es un método no destructivo que proporciona información cualitativa y cuantitativa sobre la composición química de la muestra sin generar ningún tipo de alteración o daño en ésta y no requiere el uso de reactivos (Jahurul, y otros, 2013). El uso de este tipo de información junto con algunos métodos quimiométricos como el análisis de componentes principales (PCA) y el método de mínimos cuadrados parciales (PLS), fueron utilizados en el desarrollo de este proyecto, para analizar fácilmente la información espectral permitiendo un análisis en las muestras de cacao de una forma rápida y simple (Goodacre & Anklam, 2001).

Con el fin de predecir de forma rápida el contenido de grasa se desarrolló una metodología con ventajas como: un análisis fácil sin la necesidad de un pretratamiento a las muestras antes de medir los espectros, no se necesitan reactivos, reduciendo así el costo del análisis y el tiempo análisis.

## 1. Antecedentes

En las décadas de los años 60 y 70 se empezaron a realizar investigaciones donde se implementaba la quimiometría, siendo creada en 1964 la Sociedad Internacional de Quimiometría por Wold y Kowalski (Kowalski, Brown, & Vandeginste, 1987); a partir de ese momento la química analítica se convirtió en una ciencia de información, con ayuda del desarrollo y estudio de herramientas matemáticas y estadísticas incorporadas en un software informático. Geladi y Kowalski han ejercido un importante papel en trabajos sobre PLS (Geladi & Kowalski, 1986), método matemático de gran utilidad junto con el análisis por componentes principales (PCA) que han tenido un crecimiento estable en el procesamiento de datos. Así mismo, la quimiometría en conjunto con el control estadístico multivariable de procesos (MSPC), la espectroscopía infrarroja (NIR) y diferentes métodos originados en otras disciplinas para interpretar resultados, se han implementado especialmente en la industria farmacéutica (Scafi & Pasquini, 2001), petroquímica (Macho & Larrechi, 2002) (Eide, Zahlsen, Kummernes, & Neverdal, 2006)<sup>1</sup> y alimentaria (Cozzolino, Parker, Dambergs, Herderich, & Gishen, 2006) (Craig, Franca, & Oliveira, 2012).

En particular, la espectroscopia infrarroja por Transformada de Fourier, se ha utilizado comúnmente en la industria de alimentos para el control de calidad de diferentes productos. Como en la detección de adulteraciones en vinos (Dixit, Tewari, Cho, & Irudayaraj, 2005), leche (Coitinho, y otros, 2017) y aceites comestibles (Wu, Li, & Tu, 2015), la clasificación de almidones modificados (Fernandez Pierna, Volery, Besson, Baeten, & Dardenne, 2005) y la

determinación de la contaminación de cultivos alimentarios por hongos (McMullin, Mizaikoff, & Krska, 2015). También se ha aplicado en la industria cárnica para la detección del deterioro microbiano de la carne de res (Ellis, Broadhurst, & Goodacre, 2004) y pollo (Vasconcelos, Saraiva, & De Almeida, 2014), permitiendo optimizar procesos para mejorar la vida útil de dichos productos.

En la industria del cacao, un producto de gran importancia por su consumo a nivel mundial, la quimiometría ha sido muy útil en diferentes estudios realizados. Entre estos, la clasificación de granos de cacao (Vázquez-Ovando, Molina-Freaner, Nuñez-Farfán, Betancur-Ancona, & Salvador-Figueroa, 2015) o de muestras de chocolate (Cambrai, y otros, 2010) (Cambrai, Marchioni, Julien-David, & Marcic, 2017) según su origen geográfico, la determinación y caracterización de la composición de la manteca de cacao (Maurer & Rodriguez-Saona, 2013), la estimación de forma rápida del pH y el índice de fermentación (FI) de los granos (Teye, y otros, 2015), así como la predicción rutinaria de los parámetros de calidad bioquímica basados en la cuantificación de numerosas sustancias (fenólicas, ácidos orgánicos, etc.) que representan el grado de fermentación y sabor del grano (Krähmer, y otros, 2015). Con la espectroscopía FT-IR también se han encontrado resultados de modelos prometedores para discriminar y cuantificar la adulteración en granos de cacao (es decir, mezclas de cacao fermentado y cacao no fermentado) (Teye E. , Huang, Lei, & Dai, 2014), y se ha cuantificado grasa, nitrógeno y humedad de cacao en polvo (Veselá, y otros, 2007).

Teniendo en cuenta que la determinación del contenido de grasa en cacao por métodos convencionales puede resultar relativamente tediosa y con largos tiempos de respuesta, en este trabajo se desarrolló un modelo quimiométrico predictivo para determinar el contenido total de grasa en cacao en tiempo real, de tal forma que se pueda implementar y así optimizar la

metodología utilizada actualmente para determinar esta propiedad en el control de calidad de dicho producto.

## **2. Aspectos preliminares**

### **2.1 Objetivos**

**2.1.1 Objetivo general** Desarrollar un modelo quimiométrico para predecir el contenido de grasa en granos de cacao, a partir de espectroscopia infrarroja (FTIR-ATR) y mínimos cuadrados parciales (PLS).

#### **2.1.2 Objetivos específicos**

- Determinar los parámetros espectrales óptimos para la respectiva toma de espectros de cada una de las muestras.
- Desarrollar un modelo de clasificación de las muestras.
- Obtener modelos cuantitativos para determinar el contenido de grasas en cacao.

### 3. Marco Teórico

#### 3.1 El cacao en Colombia

El árbol de cacao (*Theobroma cacao*) es una planta que crece en áreas tropicales, su cultivo es de tipo permanente, con un ciclo de duración de aproximadamente 40 años. Se conocen tres variedades morfogeográficas: criollo, forastero y trinitario. En la actualidad Colombia es uno de los países que mayor producción de cacao genera en América después de Brasil, Ecuador, Perú y República Dominicana.

Colombia cuenta con una geografía propicia para este cultivo por su ubicación, siendo Santander, Arauca, Antioquia, Huila, Nariño y Tolima los mayores productores de cacao en el país, donde el 38% de la producción es aportada por el departamento de Santander que cuenta con municipios cacaoteros por excelencia como San Vicente y Carmen de Chucurí, Landázuri, Rionegro y el Playón (Oliveros, 2017).<sup>1</sup> Según la Federación Nacional de Cacaoteros (Fedecacao), en el 2017 la producción de cacao fue de 60.535 toneladas, con un aumento del 6,6% <sup>[1]</sup> en relación con el año anterior cuando se logró una cifra histórica de 56.785 toneladas y se estima que existen cerca de 38 mil familias cultivadoras (Fedecacao, S.f.). Otro hecho importante en el 2017 está relacionado con el comercio exterior, destacándose un aumento del 11% en las exportaciones al vender 11.688 toneladas de cacao al exterior. La cadena comprende tres tipos de bienes para su comercialización: i) Primarios: cacao grano; ii) Intermedios: manteca, polvo y pasta de cacao; iii) Finales: chocolate para mesa y confites. De acuerdo con la

Organización Internacional de Cacao (ICCO) el 95% del cacao producido en Colombia se clasifica como fino de sabor y aroma, característica que posee únicamente el 5% de la producción mundial, lo que implica que el cacao colombiano es de alta calidad (Aldana, 2017). Se cree que el cacao posee beneficios para la salud, interviene en el aumento de la vasodilatación y también en la disminución de la presión arterial, lo que mejora en un gran porcentaje la salud cardiovascular y esto es gracias a que los cacaos naturales son ricos en flavonoides y poseen propiedades antioxidantes (Miller, y otros, 2008).

**3.1.1 Grasa en cacao** Las grasas y aceites vegetales son los ingredientes principales para la fabricación de varios productos alimenticios, tales como confitería, emulsiones y salsas, mantecas, margarinas, entre otros (Rios, Pessanha, Almeida, Viana, & Lannes, 2014). Las grasas y aceites son mezclas complejas conformadas por varios compuestos, donde los componentes principales son los triacilgliceroles (TAG), los diacilgliceroles (DG), los ácidos grasos libres (FFA), fosfolípidos y otros componentes menores (Van Ruth, y otros, 2010).

La manteca de cacao se utiliza principalmente en la fabricación de chocolate y sus derivados, pero también tiene un uso importante para aplicaciones en cosméticos y como un ingrediente farmacéutico. La composición física de la manteca de cacao se define empleando diferentes propiedades, tales como color, olor y sabor, el contenido de ácidos grasos libres, materia insaponificable, peróxidos, comportamiento refrigerante y humedad (Jee, 2002, págs. 66-67). En la Tabla 1 se muestran las características de una muestra típica de manteca de cacao, una muestra de buena calidad contiene entre 0,5 a 1 % de ácidos grasos libres.

Tabla 1.

*Características estándar recomendadas por la A.O.C.S (American Oil Chemists' Society) para la manteca de cacao.*

Propiedad	Valor
Densidad a 99°/15°C	0,8556-0,864
Índice de refracción a 40°C	1,453-1,458
Índice de iodo	35-40
Índice de saponificación	190-200
Materia Insaponificable %	No más de 1,0
Punto de fusión (capilar abierto), °C	28-36

Nota. Tomado de: Bailey, A. E. (1961). Aceites y grasas industriales. Reverté. Pág 114

### 3.2 Espectroscopia infrarroja

La espectroscopia infrarroja se basa en la interacción que tienen las moléculas frente a la radiación electromagnética, consiste en la medición de la radiación infrarroja la cual es absorbida por una determinada muestra (Lache-García, Meléndez-Correa, & Orrego, 2010).

La absorción de radiación en la región infrarroja es debida a la excitación de tensión la cual implica cambios en la frecuencia de vibración en los enlaces de las moléculas, y la excitación que genera deformaciones de flexión la cual implica movimiento de los átomos fuera del eje de enlace (Pasto & Johnson, 1981, pág. 122). La región IR del espectro electromagnético se divide en tres fases: el IR cercano el cual va desde 12800  $\text{cm}^{-1}$  hasta 4000  $\text{cm}^{-1}$ , el IR medio que inicia en 4000  $\text{cm}^{-1}$  hasta 200  $\text{cm}^{-1}$  y el IR lejano, desde 200  $\text{cm}^{-1}$  a 10  $\text{cm}^{-1}$ , sin embargo, el rango más utilizado está entre 4.000  $\text{cm}^{-1}$  a 670  $\text{cm}^{-1}$  (Skoog, Holler, & Timothy, 2001, págs. 409-410).

**3.2.1 Espectroscopia infrarroja con transformada de Fourier (FTIR)** En los años ochenta los instrumentos para la región del infrarrojo eran de tipo disperso, con la presencia de redes de

difracción. Actualmente existe un cambio en la instrumentación, los cuales aplican la transformada de Fourier (Skoog, Holler, & Timothy, 2001, págs. 409-410), los equipos poseen un microprocesador como parte esencial y se basan en el uso de un interferómetro y la conversión rápida de los interferogramas mediante algoritmos basados en la transformada de Fourier en espectros de absorción (Cases & Hens, 1988, pág. 739). FTIR es un método analítico no destructivo y no invasivo, requiere una cantidad mínima de muestra y ofrece una precisión bastante alta. Es una técnica multianalítica, es decir, permite realizar varias determinaciones a la vez, también es posible medir propiedades físicas y químicas, entre otras series de ventajas que posee la FTIR en comparación con otros métodos químicos (Veselá, y otros, 2007)

Por medio de la técnica FTIR es posible obtener una curva de intensidad frente al número de onda con mayor sensibilidad y de alta calidad, por esto es que es posible generar un espectro en el infrarrojo lejano, ya que anteriormente obtenerlo era complejo debido a la presencia de ruido que ocultaba la resolución (Press, 2003, pág. 295).

En el análisis de alimentos es importante el uso de la tecnología de Reflectancia Total Atenuada (ATR, por sus siglas en inglés) junto con la técnica FTIR debido a que proporciona de forma rápida información cualitativa y cuantitativa con mejor señal-ruido y una mejor resolución. El principio de la técnica se basa en la reflexión generada con el cambio de un haz de radiación que pasa de un elemento más denso (cristal ATR con alto índice de refracción) a un elemento menos denso (muestra), los cuales están en contacto; en el punto de reflexión el haz actúa como si penetrara una corta distancia de la muestra. La profundidad de penetración (dp) está dada por la ecuación (1):

$$dp = \frac{\lambda}{2\pi n_1 \left( \sin^2 \theta_i - \left( \frac{n_3}{n_2} \right)^2 \right)^{\frac{1}{2}}} \quad (1)$$

Donde  $\theta_i$  es el ángulo de incidencia,  $n_2$  es el índice de refracción del elemento ATR y  $n_3$  el Índice de refracción de la muestra (Rivero, Marder, Ferrão, Hermes, & Costa, 2008).

### 3.3 Pretratamientos espectrales

Las señales analíticas obtenidas en espectroscopía por lo general pueden presentar ruido ocasionado por diferentes razones (ruido instrumental, ruido ambiental, etc.). Para reducir o eliminar este efecto no deseado en los espectros, existen tratamientos o transformaciones que se aplican a los datos antes de ser analizados, como los que se mencionan a continuación (Romero, s.f.):

**3.3.1 Suavizado:** Esta técnica utiliza algoritmos matemáticos para reducir el ruido aleatorio sin reducir el número de variable. La mayoría de estos métodos se basan en filtros de Savitzky-Golay y en transformadas de Fourier (Romero, s.f.).

**3.3.2 Normalización:** se utiliza para obtener todos los datos en aproximadamente la misma escala. Algunos tipos de normalización son (Romero, s.f.):

- Normalización del área: Esta transformación se realiza calculando el área bajo la curva del espectro. Dicha área se vuelve igual para todas las muestras.
- Normalización del vector unitario: Los datos de una muestra  $X_i$  se normalizan a vectores unitarios.
- Normalización media: cada fila de una matriz de datos es dividida por su promedio.
- Normalización máxima: en este caso cada fila es dividida por el valor máximo absoluto en vez del promedio.

- Normalización por rango: esta normalización se da dividiendo cada fila de la matriz por su rango (valor máximo - valor mínimo).

**3.3.3 Derivación:** esta transformación trata de corregir los efectos de desviaciones de línea base en los espectros, también permite aumentar la resolución de las bandas ayudando a disminuir problemas de solapamiento de bandas. La primera derivada de un espectro es una medida de la pendiente de la curva espectral en cada punto, mientras que la segunda derivada es la medida del cambio de la pendiente en la curva (Romero, s.f.).

### 3.4 Quimiometría

La quimiometría es una técnica que se basa en la aplicación de tratamientos matemáticos y estadísticos a señales analíticas, con el fin de extraer la mayor información de los datos espectroscópicos (Rivero, Marder, Ferrão, Hermes, & Costa, 2008). Esta aproximación permite el análisis de matrices de datos que contienen muchas variables (Zuluaga, Moreno, & Quicazán, 2011), conservando únicamente la información que es relevante para describir el sistema bajo estudio (Rivero, Marder, Ferrão, Hermes, & Costa, 2008).

El uso principal de la quimiometria consiste en adquirir y aprovechar la información en las aplicaciones analíticas. El proceso de adquisición de información se basa en tres etapas donde la quimiometria está presente en cada una de diferente forma (Fernández, 2011).

- *Obtención de los datos químico-analíticos.* Al iniciar esta etapa es importante plantear el problema a resolver seguido de un muestreo, donde el tratamiento estadístico es de gran importancia. Para esta etapa es importante tener en cuenta que la información que contenga los datos sea relevante en el contexto analítico planteado.

- *Presentación condensada de la información.* Esta etapa requiere la presentación condensada de la información obtenida en forma de tablas o gráficos permitiendo así analizar la información de una forma más rápida, sencilla y clara. Es posible que se requiera el uso de algunos modelos estadísticos los cuales requieren el uso de mecanismos de cálculo.
- *Análisis estadístico de la información y conclusiones que de ella derivan.* El objetivo de esta etapa es extraer la mayor información posible de los datos analíticos (Fernández, 2011).

Se ha demostrado la eficiencia que tiene la espectroscopia infrarroja junto con la quimiometría para realizar análisis cualitativos y cuantitativos en campos como la agricultura, la alimentación, la química y la industria petrolera. La quimiometría agrupa varios temas como el diseño de experimentos, métodos de extracción de información (modelado, clasificación y prueba de la aplicabilidad de los modelos desarrollados) y técnicas que permiten entender los mecanismos químicos (Roggo, y otros, 2007).

**3.4.1 Métodos quimiométricos** Los métodos quimiométricos, son ampliamente utilizados en diferentes procesos como, en el control de la producción de cerveza y también en el control de calidad de las formulaciones farmacéuticas. Actualmente estos métodos también participan en varios procesos de fabricación, desde la pulpa y el papel, productos químicos básicos hasta productos farmacéuticos, cosméticos, bebidas y alimentos (Wold & Sjöström, 1998).

Existen diversas técnicas quimiométricas las cuales son utilizadas para el análisis de espectros obtenidos por espectroscopia infrarroja, entre las que se encuentran (Roggo, y otros, 2007):

- *Pretratamientos matemáticos.* Son utilizados con el fin de mejorar la información original que se quiere estudiar y eliminar o disminuir la información externa que puede influenciar en los espectros.
- *Métodos de clasificación.* Son utilizados para realizar la agrupación de muestras de acuerdo con su información espectral.
- *Métodos de regresión.* Son utilizados para relacionar la información espectral obtenida con propiedades cuantificables de las muestras.

Entre las técnicas mencionadas anteriormente se encuentran los métodos utilizados en el desarrollo del proyecto, el análisis de componentes principales (PCA) hace parte de los métodos de clasificación, por otro lado, el método de Mínimos Cuadrados Parciales (PLS) se puede describir como un método de regresión (Roggo, y otros, 2007).

**3.4.1.1 Análisis de componentes principales (PCA)** El análisis de componentes principales (ACP) es un método de reducción de variables donde su aplicación más importante es la reducción del número de variables no relevantes y la representación de una tabla de datos multivalente en un espacio de baja dimensión, donde las nuevas variables son combinaciones lineales de las originales y pueden interpretarse como espectros (Roggo, y otros, 2007).

Dada una matriz de datos con  $p$  variables y  $n$  muestras, donde los datos se centran primero en las medias de cada variable, asegurando así que la nube de datos está centrada en el origen de los componentes principales, sin afectar las relaciones espaciales de los datos ni las varianzas a lo largo de las variables. Los primeros componentes principales ( $Y_1$ ) están dados por la combinación lineal de las variables  $X_1, X_2, \dots, X_p$  (Roggo, y otros, 2007)

$$Y_1 = a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p \quad (2)$$

El primer componente representa la mayor varianza posible en el conjunto de datos. Se podría hacer que la varianza de  $Y_1$  fuera lo más grande posible eligiendo valores grandes para  $a_{11}$ ,  $a_{12}$ , ...  $a_{1p}$ . Para evitar esto,  $a_{11}$ ,  $a_{12}$ , ...  $a_{1p}$  se calculan teniendo en cuenta que su suma de cuadrados es igual a 1 (Holland, 2008)

$$a_{11}^2 + a_{12}^2 + \dots + a_{1p}^2 = 1 \quad (3)$$

El segundo componente principal se calcula de igual forma, teniendo en cuenta que no está correlacionado con el primer componente principal y que representa la siguiente varianza más alta (Holland, 2008).

$$Y_2 = a_{21} X_1 + a_{22} X_2 + \dots + a_{2p} X_p \quad (4)$$

Esto continúa hasta calcular  $p$  componentes principales igual al número original de variables. En este punto toda la información original ha sido explicada o contabilizada. Finalmente es posible representar las variables originales en los componentes principales como se muestra en la ecuación (5), en términos matemáticos los componentes principales ( $Y$ ) son vectores propios de la matriz de correlación (Holland, 2008).

$$Y = AX \quad (5)$$

**3.4.1.2 Método de Mínimos Cuadrados Parciales** El método PLS (Partial Least-Squares) es un algoritmo muy utilizado en quimiometría para analizar datos fuertemente colineales. Este método relaciona dos matrices de datos  $X$  (variables independientes: datos espectroscópicos) y  $Y$  (variables dependientes: propiedad a determinar) (Kumar, Bansal, Sarma, & Rawal, 2014), para obtener un número reducido de variables denominadas variables latentes, similares a los componentes principales; sin embargo, la forma como se eligen dichas combinaciones lineales

resulta diferente y se maximiza la covarianza entre las variables debido a que el objetivo de PLS es predecir las variables Y de las variables X (Miller & Miller, 2002, pág. 239).

Las matrices X e Y están dadas por las siguientes ecuaciones (Kumar, Bansal, Sarma, & Rawal, 2014):

$$X = T.P^T + E \quad (6)$$

$$Y = U.C^T + F \quad (7)$$

Donde:

T y U = matrices de puntuación, X-scores y Y-scores respectivamente.

P<sup>T</sup> y C<sup>T</sup> = matrices de carga del espacio X e Y respectivamente.

E y F = matrices individuales de X e Y (residuos)

Las variables Y se pueden predecir a partir de la matriz X-scores, de acuerdo con la Eq. 8:

$$Y = T.C^T + G \quad (8)$$

teniendo en cuenta, que T está dado por la combinación lineal de la variable X<sub>k</sub> con coeficiente de peso W\*:

$$T = X.W^* \quad (9)$$

Por lo tanto,

$$Y = X.W^*.C^T + G = XB + G \quad (10)$$

$$B = W^*.C^T \quad (11)$$

Para predecir las variables Y en muestras nuevas, se utiliza la Eq. 10, donde B es el coeficiente de PLSR y G la matriz residual (Kumar, Bansal, Sarma, & Rawal, 2014).

**3.4.2 Calibración multivariable** El objetivo de la calibración multivariable es predecir una propiedad utilizando respuestas instrumentales no selectivas con la relación entre una o más variables de respuesta “Y” y una matriz de variables predictoras “X”, como es el caso del espectro de una muestra. La variable “Y” puede ser un parámetro cualitativo o cuantitativo que representa una propiedad de interés en el sistema, y la matriz “X” contiene la información relevante de la muestra determinada por un método de análisis químico (Ferré, 2006) .

**3.4.1.3 Construcción de modelos de calibración multivariable** El proceso de modelado de varios datos tiene como finalidad obtener un modelo con el cual sea posible predecir propiedades de nuevas muestras. Para lograr un modelo firme es necesario llevar a cabo las siguientes etapas (Romero, s.f.):

- *Selección del conjunto de calibración:* En esta etapa es necesario seleccionar un conjunto con un número representativo de muestras con el que se pueda asegurar una variabilidad química y física que pueda presentarse durante el análisis. Es necesario tener en cuenta que el valor de la propiedad de interés debe conocerse para cada muestra.
- *Obtención de la señal analítica:* Las muestras deben ser sometidas al procedimiento instrumental elegido para el desarrollo del modelo, para las técnicas espectroscópicas es necesario definir las condiciones óptimas para el registro de los espectros.
- *Cálculo del modelo:* Para la construcción del modelo es necesario realizarle a la señal tratamientos previos con el fin de eliminar los efectos del ruido. Durante el desarrollo del modelo se utilizan una serie de algoritmos matemáticos y técnicas estadísticas con el fin de corroborar la calidad de este y perfeccionarlo.

- *Validación del modelo:* Esta etapa tiene como objetivo comprobar la viabilidad del modelo, es decir, si el modelo es capaz de predecir correctamente el valor de la propiedad de interés. La validación se realiza utilizando un número limitado de muestras de las cuales se conoce la propiedad, estas muestras deben ser externas al conjunto de calibración o también es posible utilizar las que participaron en la fabricación del modelo.
- *Aplicación de modelo:* En esta etapa se utilizan muestras desconocidas y tiene como finalidad analizar los resultados, esta fase se podría describir como una segunda comprobación de la capacidad predictiva del modelo (Romero, s.f.).

#### **4. Metodología**

La metodología del proyecto se llevó a cabo siguiendo el esquema que se ilustra en la figura 1, donde cada etapa se realizó de forma consecutiva, con el fin de cumplir con los objetivos planteados en el trabajo. El proyecto se desarrolló con el apoyo y orientación del Laboratorio de Espectroscopia Atómica y Molecular (LEAM), junto con la colaboración del Grupo de Investigación en Ciencia y Tecnología de Alimentos (CICTA).

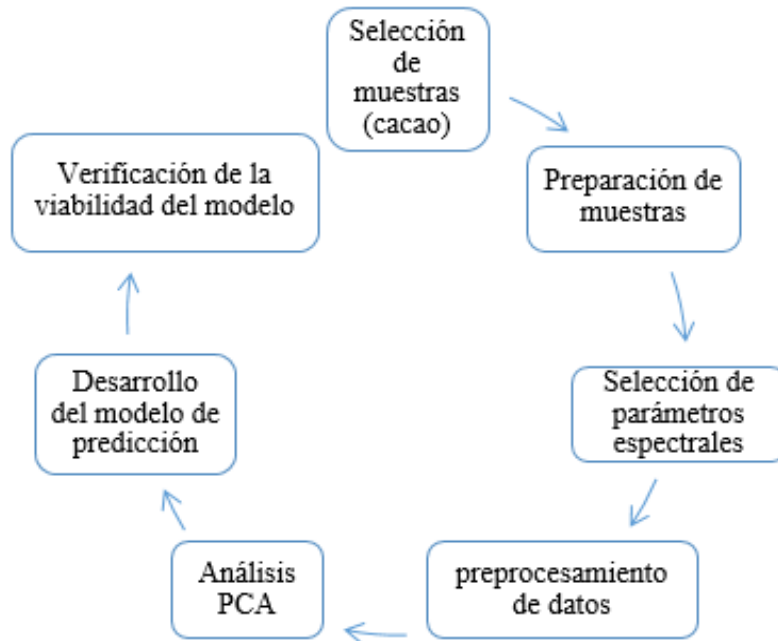


Figura 1. Esquema de la metodología llevada a cabo.

#### 4.1 Materiales y reactivos

**4.1.1 Reactivos** Una de las ventajas de esta metodología es que no se necesitan reactivos, durante su desarrollo sólo se empleó alcohol, utilizado para la limpieza del equipo en cada toma de espectro.

**4.1.2 Equipos** La colección de los espectros NIR se llevó a cabo en un Espectrómetro Infrarrojo con transformada de Fourier, modelo Thermo Scientific <sup>TM</sup> iS<sup>TM</sup> 50 FT-IR, que se observa en la Figura 2, el equipo se encuentra en el Laboratorio de Espectroscopia Atómica y Molecular (LEAM) de la Universidad Industrial de Santander.



*Figura 2.* Espectrómetro Infrarrojo utilizado en el desarrollo del proyecto.

#### **4.2 Selección de muestras**

Para el desarrollo de métodos quimiométricos es necesario contar con un número representativo de muestras, de las cuales se debe conocer el valor de la propiedad de interés. En este caso fueron utilizadas 33 muestras de cacao colombiano fermentado y seco de distintas regiones como Santander, Villeta, Boyacá y Mosquera, suministradas por el Grupo de Investigación en Ciencia y Tecnología de Alimentos (CICTA) de la Universidad Industrial de Santander, así como los respectivos valores de su contenido de grasa (obtenidos previamente por el método AOAC 963.15, Fat in Cacao Products), los cuales se observan en la tabla 2.

Tabla 2.

*Porcentaje de grasa de treinta y tres muestras de cacao colombianos.*

<b>N° de Muestra</b>	<b>Origen de las Muestras</b>	<b>Primera Medición</b>	<b>Segunda Medición</b>	<b>% p/p Promedio de Grasa en Base Seca</b>
51	Desconocido	52,5952	51,5432	52,07
52	Desconocido	47,6497	48,3109	47,98
53	Desconocido	47,8455	47,3602	47,60
54	Desconocido	46,9681	47,2216	47,09
55	Desconocido	51,2851	51,0288	51,16
56	Desconocido	49,0330	49,1805	49,11
57	Desconocido	45,5273	45,3653	45,45
58	Desconocido	48,5351	47,2103	47,87
59	Desconocido	50,8049	49,4392	50,12
60	Desconocido	50,6884	49,4996	50,09
61	Desconocido	53,3507	52,5840	52,97
62	Desconocido	51,0759	51,5057	51,29
63	Desconocido	45,3390	45,2596	45,30
79	Santander	53,0486	52,4835	52,77
82	Santander	55,4423	54,5007	54,97
83	Santander	54,0797	54,8422	54,46
86	Santander	54,1212	53,7708	53,95
89	Santander	54,0291	54,4805	54,25
92	Santander	54,3366	53,7881	54,06
186	Villeta Cundinamarca	54,1932	53,2959	53,74
187	Villeta Cundinamarca	54,2734	54,0061	54,14
188	Villeta Cundinamarca	53,4867	53,6718	53,58
189	Villeta Cundinamarca	54,2530	53,2709	53,76
190	Villeta Cundinamarca	52,8326	51,9899	52,41
191	Villeta Cundinamarca	54,2833	53,3779	53,83
192	Villeta Cundinamarca	51,2795	51,7117	51,50
194	Villeta Cundinamarca	51,3897	51,7340	51,56
195	Villeta Cundinamarca	53,0633	56,2327	54,65
750	Boyacá	55,1070	55,0270	55,07
751	Boyacá	55,8457	55,0598	55,45
822	Mosquera Cundinamarca	56,3397	55,7027	56,02
823	Mosquera Cundinamarca	54,6620	54,6958	54,68

N° de Muestra	Origen de las Muestras	Primera Medición	Segunda Medición	% p/p Promedio de Grasa en Base Seca
824	Mosquera Cundinamarca	55,6577	54,3619	55,01

### 4.3 Preparación de muestras para la toma de espectros

Para la finalidad del estudio las muestras se sometieron a una molienda para reducir y homogenizar el tamaño de partícula. Los espectros fueron tomados a temperatura ambiente en un rango de 400 a 4000 ( $cm^{-1}$ ), puesto que en esta región se puede detectar la presencia de los grupos funcionales que indicaran la existencia de grasa de cacao tales como los grupos OH y los grupos CO, que indican la presencia de ácidos grasos o ésteres que conforman a los glicéridos componentes de la grasa de cacao. El aire fue tomado como espectro de referencia.

### 4.4 Optimización de las condiciones experimentales

Se realizó un análisis preliminar para la toma de los espectros de las muestras utilizando las condiciones establecidas por el equipo de infrarrojo correspondientes a 32 barridos, una velocidad de 0,4747 cm/s y una resolución de 4  $cm^{-1}$ . Para mejorar la calidad de los espectros y establecer las condiciones experimentales óptimas, se realizó una variación de dichos parámetros y se escogieron aquellos con los cuales se obtuvieron espectros más limpios, es decir con menor ruido y una mayor relación señal/ruido.

#### 4.5 Pretratamiento de los espectros

En esta etapa se aplicó un procesamiento a los datos espectroscópicos para luego hacer el análisis quimiométrico, dicho procesamiento al igual que el análisis de PCA se realizaron utilizando el software Unscrambler 10.3, mientras que el modelo de regresión por PLS se realizó bajo un código de Matlab.

**4.5.1 Suavizado** Se realizó un suavizado espectral para reducir el ruido aleatorio existente en la señal instrumental. Se utilizó el suavizado Gaussian Filter que ofreció un resultado más favorable en comparación con los otros métodos de suavizado que brinda el software Unscrambler, generando una disminución del ruido sin reducir el número de variables.

**4.5.2 Normalización.** La normalización fue el segundo tratamiento que se le aplicó a los espectros de las muestras de granos de cacao con el fin de obtener todos los datos en la misma escala, para el análisis de componentes principales (PCA) se utilizó la normalización por área.

Para el desarrollo del modelo mediante el método de Mínimos Cuadrados Parciales (PLS) se utilizó la normalización de la variación sistemática (SVN), con el fin de aumentar la relación señal-ruido, esta consiste en que a cada espectro se le subtrae la media calculada a partir de los espectros de las 33 muestras y luego se divide sobre la desviación estándar.

**4.5.3 Primera derivada.** Se calculó la primera derivada de los espectros debido a que la primera derivada es más sensible a las variaciones de intensidad en el espectro, lo que permite

establecer diferencias en los datos espectrales, como solapamiento en los picos. En este caso utilizó la derivada de Savitzky-Golay.

#### **4.6 Análisis por componentes principales.**

Seguido al preprocesamiento, por medio del software Unscrambler 10.3 se realizó un análisis PCA mediante la descomposición de los datos espectrales en componentes principales. Se realizó este análisis con el fin de discriminar las muestras por sus parecidos o sus diferencias.

#### **4.7 Construcción del modelo**

La correlación entre el contenido de grasa y los espectros NIR de las muestras de cacao se realizó con el método de regresión por mínimos cuadrados parciales (Partial Least Squares, PLS) implementando el código de Matlab, mediante una etapa de calibración y otra de validación, implícitas en este método cuantitativo. Se utilizó el método de validación cruzada (CV-Cross Validation) de Montecarlo, dejando un 60% de las muestras para la calibración del modelo y el 40% restante para la validación. Para esto se utilizaron dos matrices: la matriz de datos predictores (X) proveniente de los espectros y la matriz de respuesta (Y) correspondiente al porcentaje de grasa. Se construyeron dos modelos aceptables, el primero aplicando una normalización por área y el segundo una normalización por z-score como se observa en la sección de resultados.

#### 4.8 Verificación de la viabilidad del modelo

La aplicabilidad del método se evaluó utilizando parámetros estadísticos como el coeficiente de regresión lineal de los modelos y el error cuadrático medio de calibración y validación, RMSEC y RMSEV.

### 5. Resultados

#### 5.1 Parámetros instrumentales

**5.1.1 Variación del número de SCAN.** Para este procedimiento se utilizó la muestra 63, la cual se escogió de forma aleatoria entre el grupo de muestras utilizadas para el desarrollo del proyecto, se tomaron los espectros con diferente número de SCAN (figura 3) dejando fijos los valores de la velocidad del espejo y la resolución. Teniendo en cuenta que las señales obtenidas en los espectros están dadas en % de Transmitancia, los datos fueron pasados a Absorbancia de acuerdo con la ecuación 12 (Mackenzie, 2005) y a partir de estos se determinó la relación señal/ruido para el pico presente a  $2914,925 \text{ cm}^{-1}$  como se muestra en la tabla 3.

$$A = 2 - \text{Log}(\% \text{ Transmitancia}) \quad (12)$$

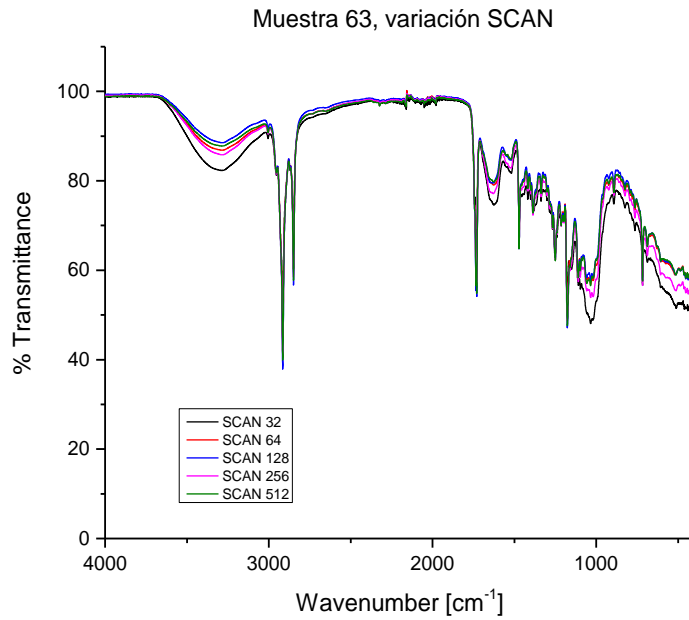


Figura 3. Espectros de la muestra 63, tomados con diferente número de SCAN.

Tabla 3.

Relación Señal/Ruido obtenidas con los espectros tomados a diferentes números de SCAN.

Número de SCAN	Señal (cm <sup>-1</sup> )		Ruido (cm <sup>-1</sup> )			S/R
	2914,925	2990,619	2981,941	2792,948	2736,539	
<b>Absorbancias</b>						
<b>32</b>	0,31939	0,04371	0,04629	0,03108	0,02606	2,17069
<b>64</b>	0,36170	0,03701	0,03940	0,02670	0,02211	2,88885
<b>128</b>	0,42072	0,03303	0,03553	0,02346	0,01890	3,79270
<b>256</b>	0,39051	0,03819	0,04084	0,02728	0,02244	3,03314
<b>512</b>	0,39825	0,03656	0,03894	0,02707	0,02240	3,18652

De acuerdo con los resultados registrados en la tabla 3, se estableció que el número de SCAN óptimo para la toma de espectros es de 128, ya que con este número de barridos se obtuvo una mayor relación S/R que con los otros valores utilizados.

**5.1.2 Variación de la resolución espectral.** Una vez escogido el número de SCAN óptimo, se dejó fijo este parámetro junto con la velocidad del espejo establecida por el equipo y se varió la resolución, utilizando dos valores  $4\text{ cm}^{-1}$  y  $2\text{ cm}^{-1}$ . Los espectros obtenidos se encuentran en la figura 4.

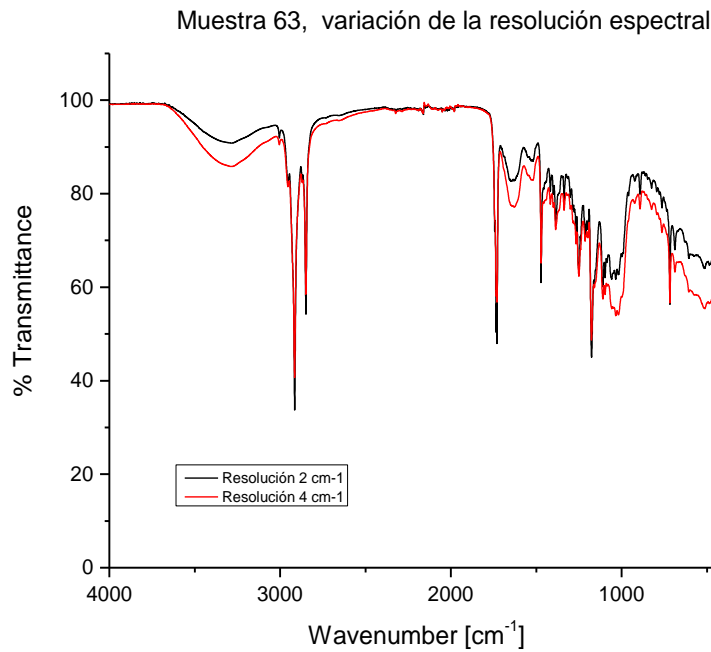


Figura 4. Espectros de la muestra 63, tomados con diferente resolución espectral.

Posteriormente se calculó la relación señal/ruido (tabla 4) para la misma señal utilizada en la selección del número de SCAN ( $2914,925\text{ cm}^{-1}$ ) y se determinó que la resolución adecuada para trabajar es de  $2\text{ cm}^{-1}$ .

Tabla 4.

Relación Señal/Ruido obtenidas con los espectros tomados a diferente resolución espectral.

Resolución espectral (cm <sup>-1</sup> )	Señal (cm <sup>-1</sup> )		Ruido (cm <sup>-1</sup> )			S/R
	2914,925	2990,619	2981,941	2792,948	2736,539	
	Absorbancias					
2	0,46926	0,02917	0,03135	0,02097	0,01676	4,77582
4	0,39051	0,03819	0,04084	0,02728	0,02244	3,03314

**5.1.3 Variación de la velocidad del espejo.** Dejando fijos el número de SCAN y la resolución espectral establecidos anteriormente, se tomaron los espectros a diferentes velocidades del espejo (figura 5) y se determinó la relación señal/ruido en cada uno de ellos como se observa en la tabla 5, escogiendo de este modo una velocidad optima de 0,4747 cm/s.

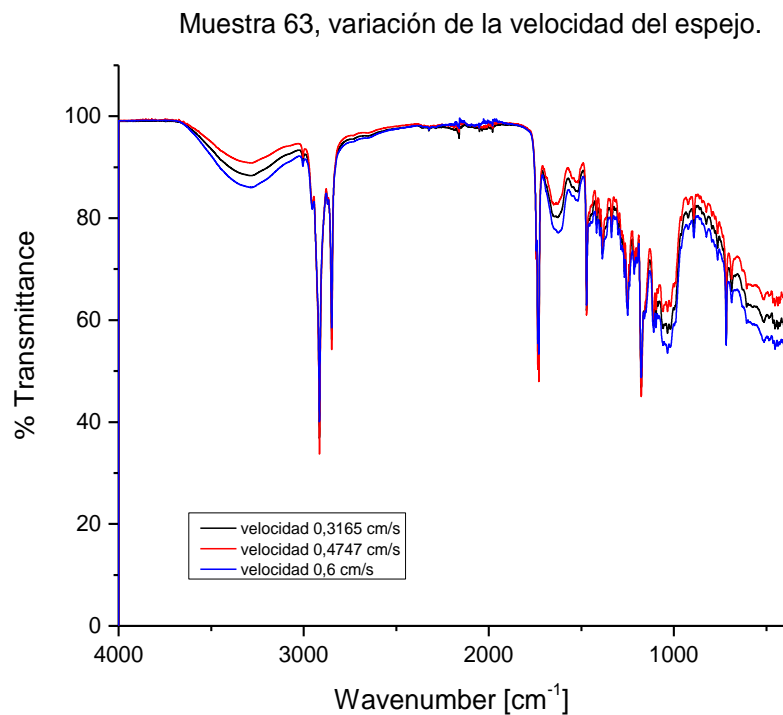


Figura 5. Espectros de la muestra 63 tomados con diferentes valores de velocidad del espejo.

Tabla 5.

*Relación Señal/Ruido obtenidas con los espectros tomados a diferente velocidad del espejo.*

Velocidad del espejo (cm/s)	Señal (cm <sup>-1</sup> )		Ruido (cm <sup>-1</sup> )			S/R
	Absorbancias					
	2914,925	2990,619	2981,941	2792,948	2736,539	
<b>0,3145</b>	0,43281	0,03386	0,03617	0,02439	0,01992	3,78565
<b>0,4747</b>	0,46926	0,02917	0,03135	0,02097	0,01676	4,77582
<b>0,6</b>	0,39661	0,03838	0,04048	0,02720	0,02246	3,08615

Finalmente, por medio de la variación señal-ruido y el análisis de los espectros mostrados en las figuras 3, 4 y 5 fue posible determinar los parámetros espectrales óptimos para la toma de espectros los cuales se observan en la tabla 6.

Tabla 6.

*Parámetros establecidos para la toma de espectros de cada una de las muestras.*

Parámetro	Valor
Número De SCAN	128
Resolución Espectral	2 cm <sup>-1</sup>
Velocidad Del Espejo	0,4747 cm/s

**5.1.4 Verificación del rendimiento del equipo.** En esta etapa se midió la incertidumbre o precisión de los datos espectrales, con el fin de verificar el desempeño del espectrofotómetro, para ello se realizó una prueba de repetibilidad, tomando 20 veces el espectro de la muestra 56 (figura 6) la cual fue escogida en forma aleatoria con la toma de espectros de otras muestras y se calculó la desviación estándar de la variación de la intensidad de las señales presentes a 2914,68 cm<sup>-1</sup> y 1743,12 cm<sup>-1</sup>. Como se muestra en la tabla 7, se obtuvieron desviaciones de 2,783 y 0,947 respectivamente para cada señal, lo que nos permite garantizar la calidad de los datos y la fiabilidad del equipo.

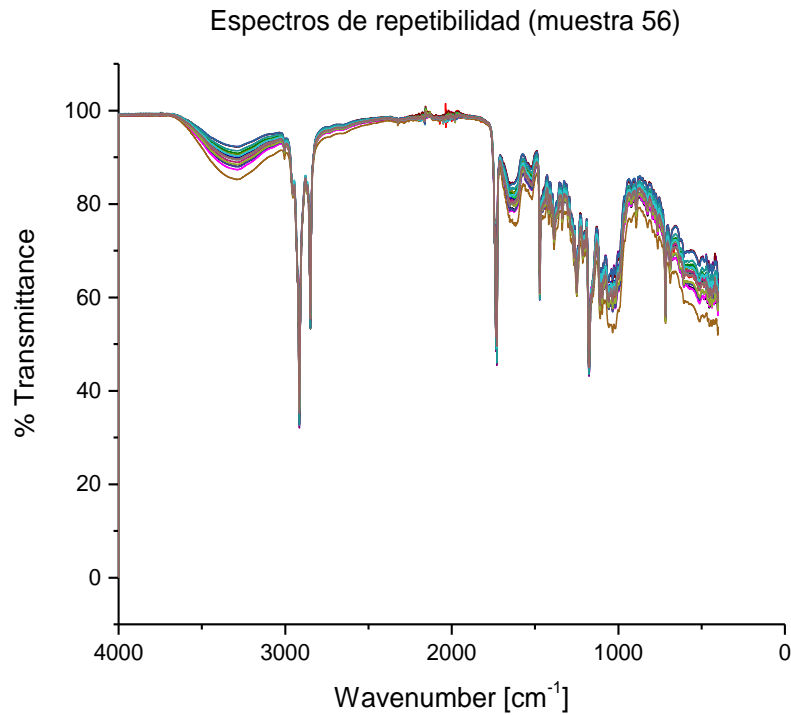


Figura 6. Espectros de la muestra 56 obtenidos en la prueba de repetibilidad.

Tabla 7.

Desviación estándar del % de transmitancia a 2915,68 cm<sup>-1</sup> y 1743 cm<sup>-1</sup> de las 20 repeticiones.

No. Repetición	% de Transmitancia a:	
	2914,68 cm <sup>-1</sup>	1743,12 cm <sup>-1</sup>
1	39,5636	73,1850
2	34,5472	72,2456
3	36,0199	72,4951
4	40,7238	74,0052
5	32,9740	71,8630
6	34,1530	73,1554
7	32,6582	72,2172
8	32,1293	72,0703
9	32,9617	71,5446
10	33,8910	71,7994
11	33,1098	71,3617
12	32,6753	72,1422
13	41,1150	74,8417

No. Repetición	% de Transmitancia a:	
	2914,68 cm <sup>-1</sup>	1743,12 cm <sup>-1</sup>
14	36,8825	73,8685
15	38,7397	73,7042
16	35,8158	73,9848
17	37,9136	73,3358
18	36,1935	72,7927
19	37,0087	72,3083
20	35,1935	72,9196
<b>Desviación estándar</b>	2,78280	0,94674

### 5.2 Pretratamiento espectral

Para mejorar la relación señal/ruido en las señales analíticas de las muestras de cacao fue utilizado el suavizado Gaussian Filter. En la figura 7 se evidencian los espectros sin pretratamiento alguno de las 33 muestras utilizadas en el desarrollo del proyecto, comparándolos con los espectros que se muestran en la figura 8 a los cuales se les aplicó el método de suavizado es posible observar espectros con menos ruido y más claros.

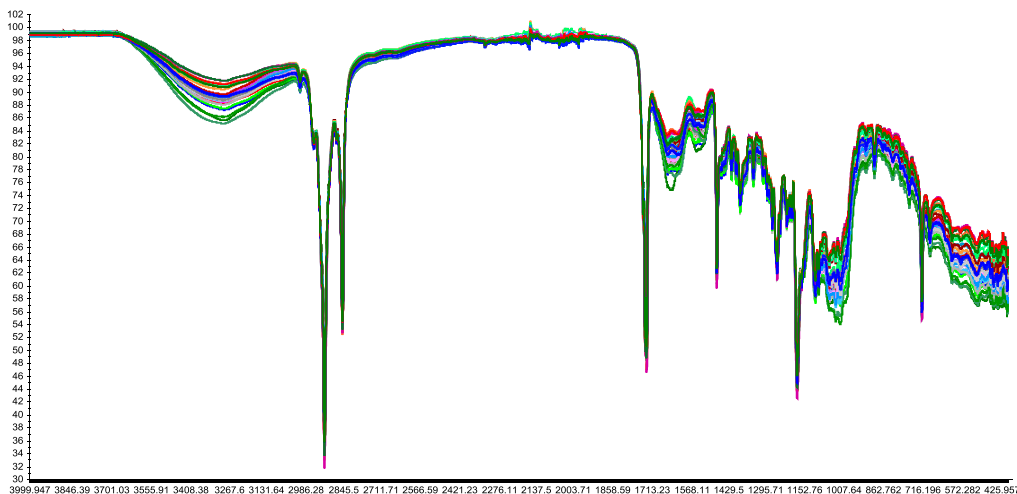


Figura 7. Espectros de las 33 muestras sin pretratamiento alguno.

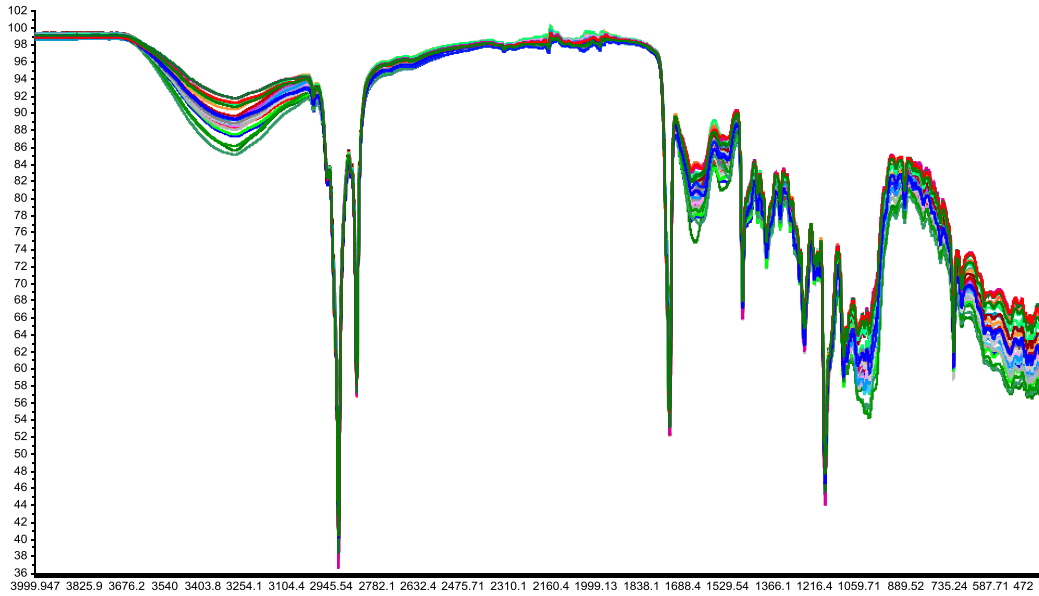


Figura 8. Espectros de las 33 muestras con el método de suavizado Gaussian Filter.

Por otro lado, en el análisis de componentes principales (PCA) se aplicó una normalización por área, debido a que esta transformación brindó espectros más limpios (figura 9) en comparación con los otros tipos de normalización que ofrece el software Unscrambler 10.3. Así mismo en el modelo predictivo desarrollado por PLS fue utilizada la normalización de la variación sistemática (SVN), en la figura 10 se evidencian los espectros normalizados con SVN.

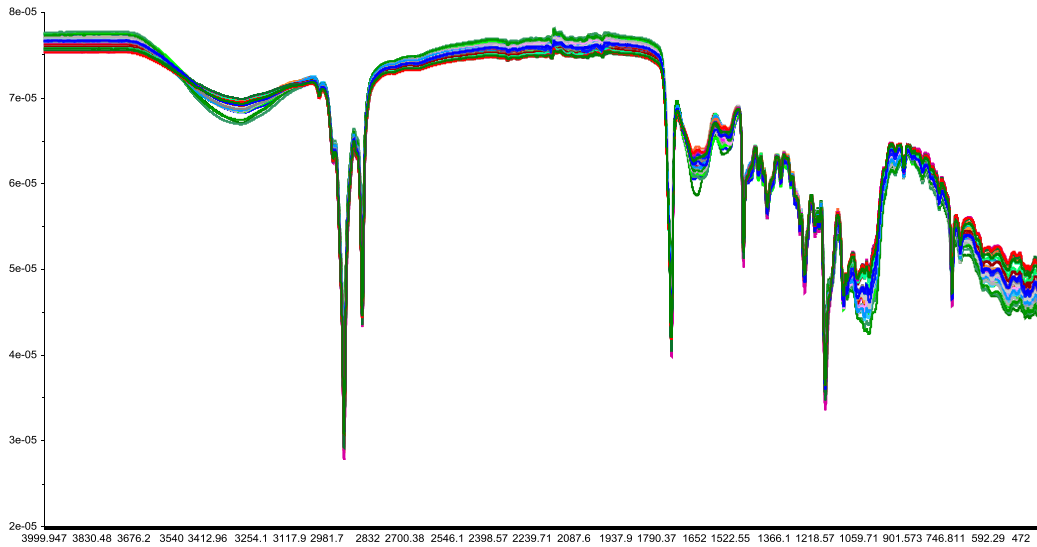


Figura 9. Espectros de las 33 muestras normalizados.

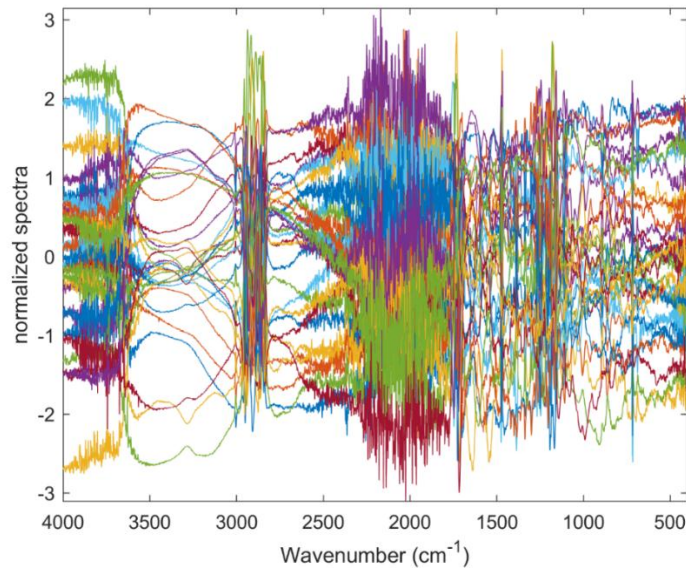


Figura 10. Espectros de las 33 muestras tratados con normalización SVN.

Finalmente se calculó la primera derivada, aplicando la derivada de Savitzky-Golay que consiste en realizar un ajuste de regresión lineal de mínimos cuadrados de un polinomio alrededor de cada punto del espectro para suavizar los datos; por lo tanto, la derivada es entonces

la derivada del polinomio ajustado en cada punto. El resultado de aplicarle este tratamiento a los espectros de las muestras se puede evidenciar en la figura 11.

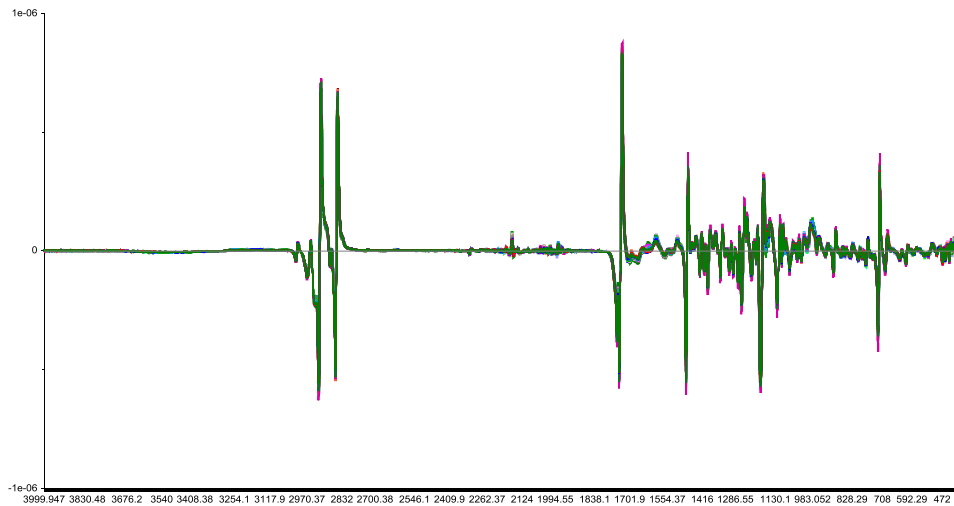


Figura 11. Derivada Savitzky-Golay aplicada a los espectros de las 33 muestras.

### 5.3 Análisis por componentes principales (PCA).

A continuación, se presentan los resultados generados en el software Unscrambler 10.3 al realizar el Análisis por Componentes Principales (PCA). La matriz de datos utilizada en este análisis está conformada por 33 filas las cuales corresponden al número de muestras y 14935 columnas que pertenecen a las señales de los espectros entre 400 y 4000  $\text{cm}^{-1}$  para cada muestra.

En la Figura 12 se observa un diagrama de dispersión para dos componentes, PC1 y PC2, por medio de esta dispersión es posible asumir que cuanto más cerca están las muestras en el gráfico, más similares son respecto a los dos componentes en cuestión, es decir que este análisis brinda información sobre la similitud o diferencia entre las muestras. En esta gráfica de dispersión se observa que la mayoría de las muestras se agrupan de acuerdo con su lugar de origen,

identificando así cuatro grupos de muestras correspondientes a: Villeta Cundinamarca (elipse azul), Santander (elipse morada), Mosquera Cundinamarca (elipse verde) y finalmente en la parte superior del diagrama se observa la elipse más grande resaltada en rojo que corresponde a un grupo de muestras de origen desconocido, a pesar de esto es posible observar que estas muestras tienden agruparse en un mismo grupo.

Por medio del diagrama de dispersión PC1 y PC2 también fue posible realizar una aproximación que permite dividir las muestras en dos grandes grupos (figura 13), el primero en la parte superior (primer y segundo cuadrante de la gráfica - elipse roja) el cual corresponde a las muestras que poseen el menor porcentaje de grasa, por el contrario, las muestras con mayor porcentaje de grasa están representadas en la parte inferior del diagrama (tercer y cuarto cuadrante de la gráfica) por una elipse de color amarillo.

En la figura 14 se puede observar la elipse  $T^2$  de Hotelling, aplicada al diagrama de dispersión de las 33 muestras, aplicando el método de  $T^2$  Hotelling es posible identificar muestras atípicas, en este caso ninguna muestra se encuentra fuera del límite de confianza el cual equivale al 95% de modo que no se presentan muestras atípicas (outliers).

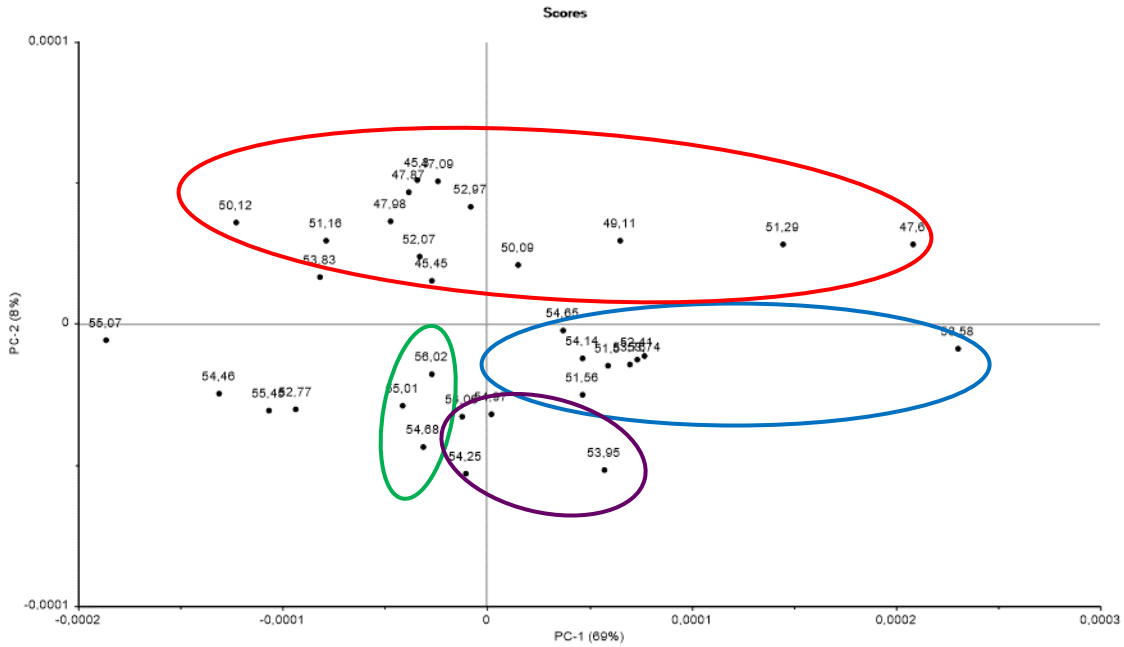


Figura 12. Agrupación según el origen de las 33 muestras de cacao en la gráfica de Scores para el PC1 y PC2.

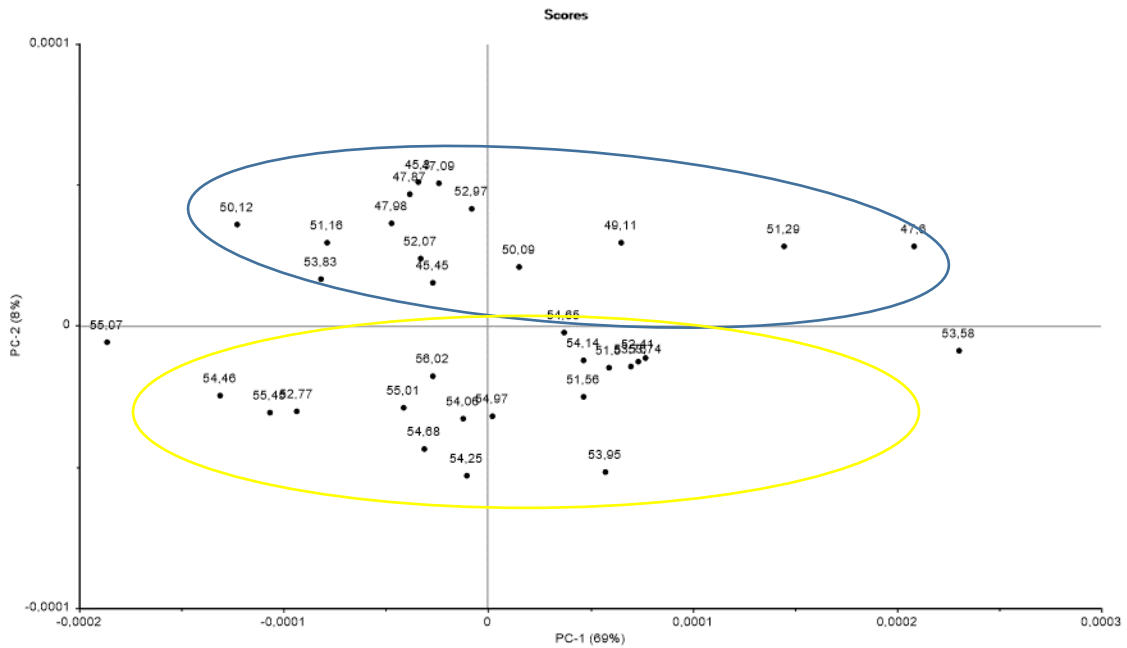


Figura 13. Agrupación según el porcentaje de grasa de las 33 muestras de cacao en la gráfica de Scores para el PC1 y PC2.

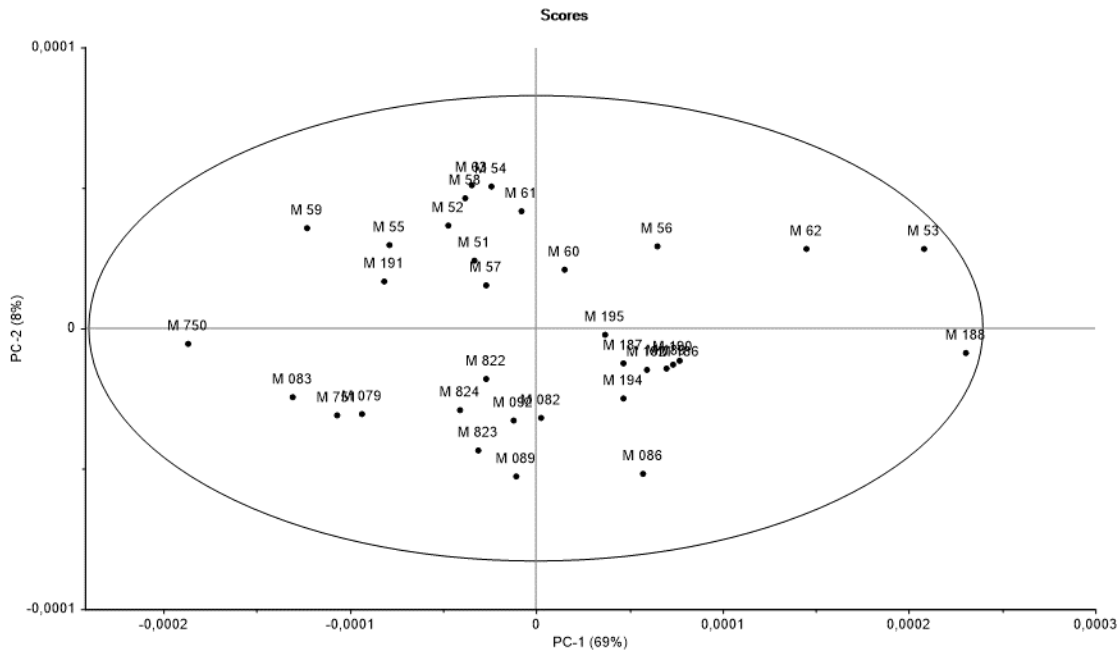


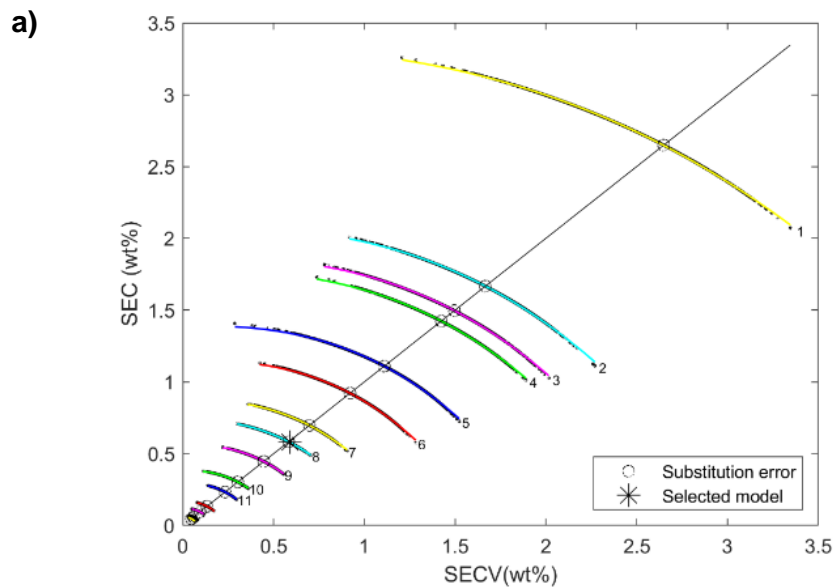
Figura 14.  $T^2$  de Hotelling aplicado a la gráfica de Scores de las 33 muestras de cacao.

### 5.4 Método de mínimos cuadrados parciales (PLS)

A partir del análisis de los datos espectroscópicos junto con el algoritmo NIPALS (Nonlinear Iterative Partial Least Square) trabajado en Matlab y el método matemático PLS se propusieron dos modelos predictivos variando el tipo de normalización. En el código de Matlab utilizado, se aseguró que al calcular dichos modelos el valor mínimo (45,30%) y máximo (56,02%) de la propiedad estén en los extremos cerrando el conjunto de datos de la calibración para garantizar la predicción del contenido de grasa dentro de este rango.

**5.4.1 Determinación del número de componentes o variables latentes** El primer modelo propuesto fue desarrollado aplicando una normalización por área mientras que el segundo una normalización z-score. En los dos casos la complejidad o número de componentes utilizados fue

de ocho, este número se escogió a partir del análisis del error cuadrático medio de predicción (RMSEP). Para esto se tuvo en cuenta la gráfica del (RMSEP) contra el (RMSECal) para cada modelo obtenido, que corresponde a cada punto observado en el espacio coordinado de dicha gráfica (figura 15 a-b), allí se puede ver el comportamiento de estos dos errores para 1000 modelos generados con diferente número de componentes. Los puntos de corte con la recta identidad encerrados en círculos concuerdan con el error de sustitución (ocasionado al entrenar y validar con todos los datos). Es importante que el modelo seleccionado tenga un optimismo moderado y una complejidad moderada, es decir, un modelo que no tenga demasiadas componentes que produzcan un sobreajuste (lo cual implica la memorización del modelo), o tan pocas componentes que se obtenga un modelo subestimado. Por esta razón los modelos seleccionados se encuentran dentro del conjunto de modelos de ocho componentes como se observa en la figura 15.



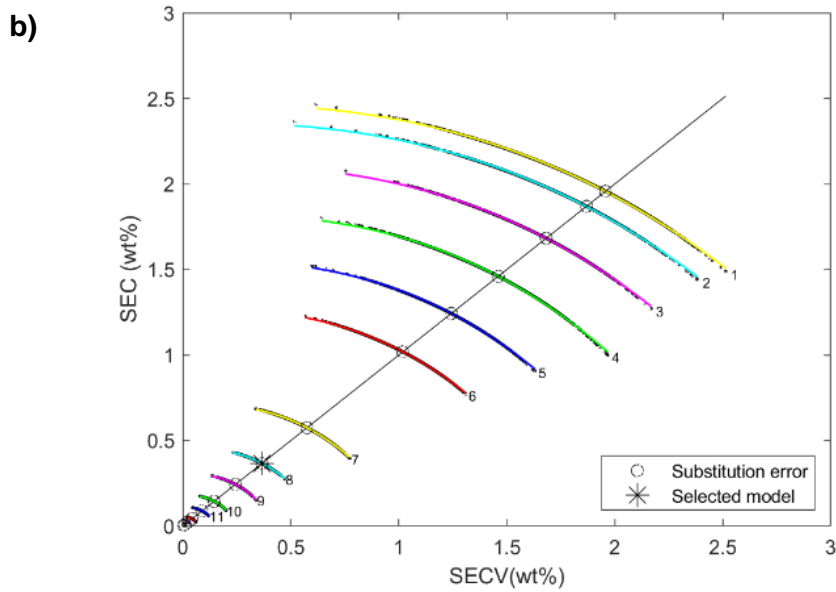


Figura 15. Curva de tendencia de RMSEP contra RMSECal de predicción del porcentaje de grasa en cacao con el pretratamiento de: a) Normalización por área, b) Normalización z-score.

Otra estrategia utilizada para determinar el número de componentes fue la gráfica de variables latentes vs el error (cociente de la dispersión de los errores entre los errores de sustitución) (figura 16 a-b), tomando el mínimo local que corresponde al mínimo, en el caso del modelo de normalización por área este valor corresponde a 13 componentes (figura 16-a), lo que se considera demasiadas componentes puesto que conllevaría a un modelo sobreajustado, por esto se estableció que el número de componentes indicado es ocho, que corresponde al primer mínimo local.

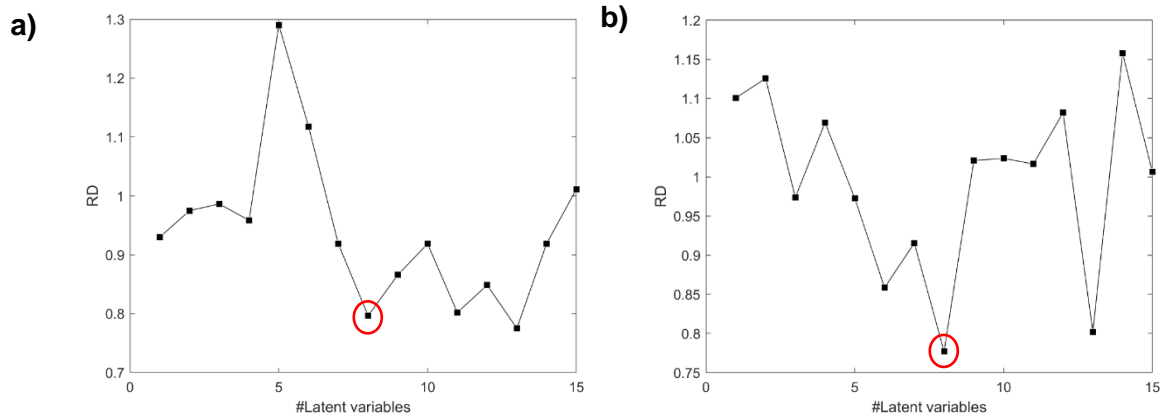


Figura 16. (RMSEP) contra el número de componentes para los modelos con pretratamiento de: a) Normalización por área, b) Normalización z-score.

**5.4.2 Modelos propuestos** Se desarrollaron dos modelos predictivos para determinar el contenido de grasa en cacao fermentado y seco, el primer modelo se realizó a partir de una normalización por área, y el segundo a partir de una normalización por z-score. En los dos modelos se observa una alta correlación entre los datos predichos y los reales, lográndose coeficientes de regresión de 0,96 y 0,98; errores estándar de la calibración (SEC) de 0,57 y 0,36 y errores estándar de la validación cruzada (SECV) de 0,58 y 0,36, para la normalización por área y z-score respectivamente. Sin embargo, el modelo por normalización z-score posee un rendimiento considerablemente mejor que el modelo a partir de normalización por área, con el cual no se obtiene una buena distribución de los datos de validación y calibración en el rango de 46% a 50% del eje x (contenido de grasa medido). Los datos de calibración son bajos en comparación con los datos de validación, debido a que las muestras se escogieron aleatoriamente por la validación cruzada de Montecarlo tal como se explicó en la sección de la metodología, sin embargo, los valores de validación caen dentro del rango del 95% de confianza. En la Figura 17 se observa la curva de calibración para cada modelo, donde se evidencia que la distribución de los datos es mejor para el modelo realizado bajo la normalización z-score.

Por otro lado, se analizaron los residuales que son equivalentes a la diferencia entre el valor de la propiedad predicho por el modelo y el valor real de propiedad, de forma que los residuos describen el error aleatorio del modelo de regresión, el análisis del comportamiento de estos es indispensable para formular un diagnóstico sobre la exactitud del modelo. Como se observa en la *Figura 18* los residuales no presentan ningún sesgo, es decir no tienen una tendencia, lo cual muestra linealidad y asegura que los modelos no presentarán un resultado de memorización para las muestras utilizadas en su desarrollo y errores de predicción en muestras externas e independientes a éste. En esta figura también se observa que el modelo obtenido utilizando normalización z-score se ajusta mejor a los datos como lo demuestra los residuos más pequeños, en comparación con el modelo por normalización por área.

Además, la gráfica de los histogramas de los residuales para el modelo por z-scores (*figura 19-a*) muestra una tendencia de probabilidad normal, donde la mayor probabilidad se encuentra en la parte central del histograma alrededor de cero.

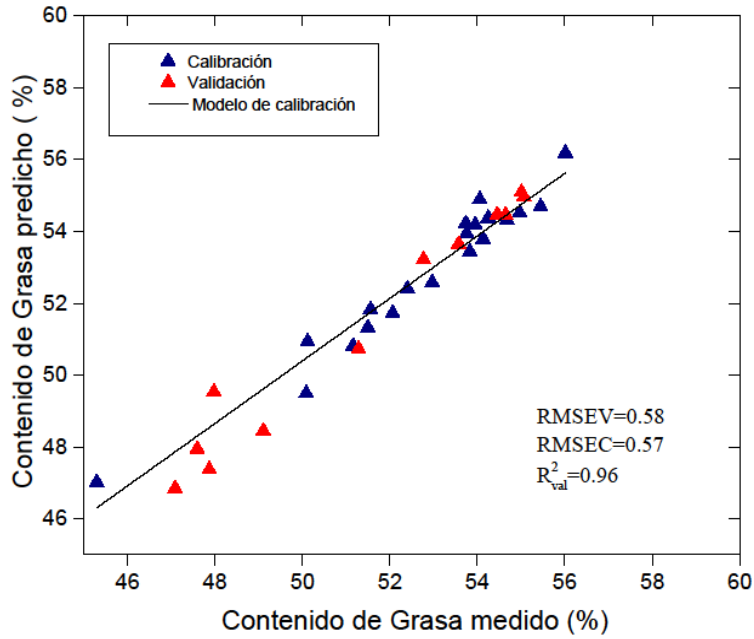
En las Tablas 8 y 9 se observan los valores de los porcentajes de grasa obtenidos experimentalmente y los valores de los porcentajes predichos por cada modelo para el set de muestras utilizadas en la validación. Al comparar los valores predichos es posible corroborar que los mejores resultados se obtienen con el modelo por z-score, donde el error cuadrático medio de predicción es menor.

En estudios anteriores (Veselá, y otros, 2007) utilizaron la técnica de Infrarrojo con Transformada de Fourier (FTIR) para construir modelos de calibración para la determinación del contenido de grasa en cacao, y obtuvieron un modelo predictivo con un RMSECV relativo de 10.4% y un  $R^2 = 0.94$ , dichos autores trabajaron con muestras de cacao en polvo comercial y para realizar la cuantificación de la grasa utilizaron la segunda derivada de los espectros, sin

embargo, con la metodología propuesta en este proyecto fue posible desarrollar dos modelos que predicen el contenido de grasa en muestras de cacao fermentado en base seca con errores de predicción de 0,36 y 0,58%.

En Colombia no se evidencian estudios en los que se utilice la espectroscopía FT-IR junto con métodos quimiométricos y la implementación de la validación cruzada de Montecarlo para la determinación del porcentaje de grasa en muestras de cacao, ya que normalmente se utiliza la validación cruzada “leave-one-out”, donde existe una probabilidad bastante alta de un ajuste excesivo, y por lo tanto, una subestimación del verdadero error predictivo (Xu, 2001). Es por esto que la metodología descrita en este proyecto provee soluciones a los tediosos métodos de determinación de grasa en cacao.

a)



b)

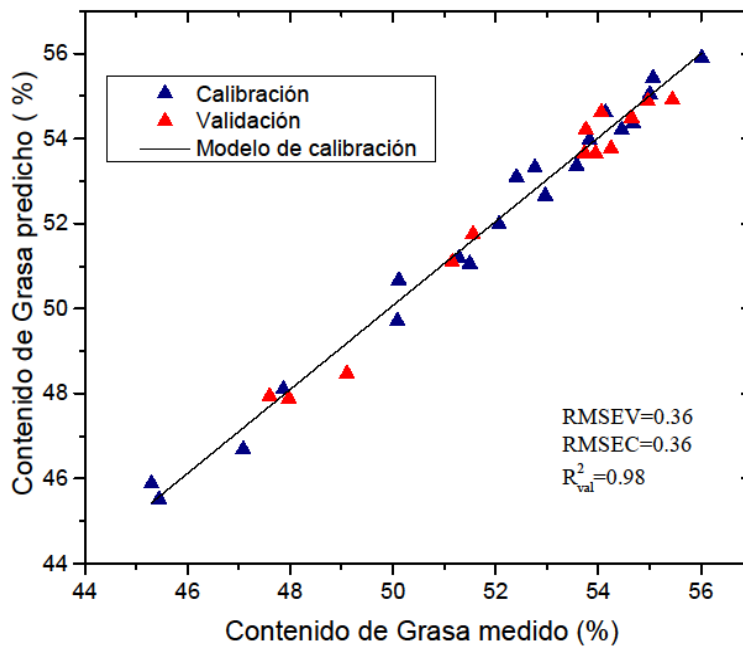


Figura 17. Curva de calibración para el modelo de predicción de grasa de cacao bajo un pretratamiento de: a) Normalización por área, b) Normalización z-score.

Tabla 8.

*Valores medidos y predichos para el set de muestras de validación en el modelo desarrollado a partir de normalización por área. Error cuadrático medio de predicción de 0,36*

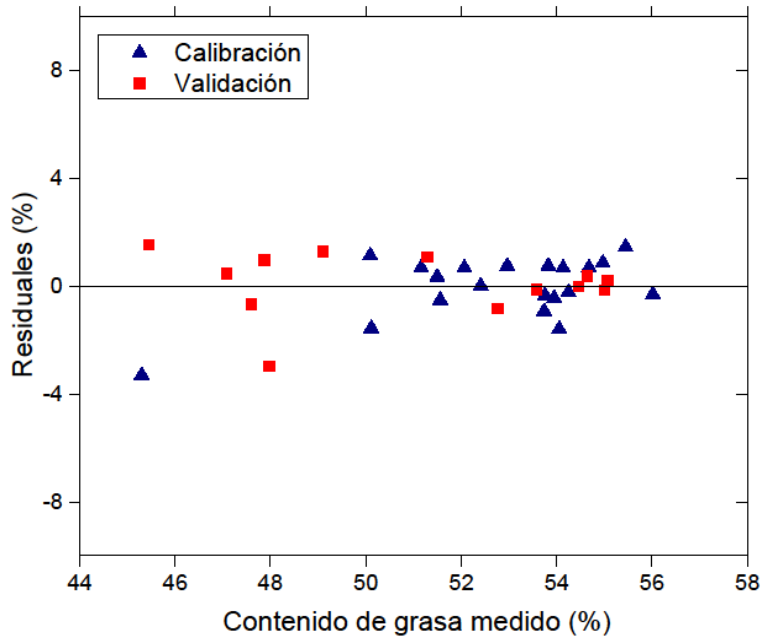
<b>Muestra</b>	<b>% Grasa experimental</b>	<b>%Grasa predicho</b>
188	53,58	53,65
195	54,65	54,46
52	47,98	49,53
53	47,6	47,94
54	47,09	46,85
56	49,11	48,44
57	45,45	44,64
58	47,87	47,37
62	51,29	50,73
79	52,77	53,21
83	54,46	54,46
750	55,07	54,96
824	55,01	55,09

Tabla 9

*Valores medidos y predichos para el set de muestras de validación en el modelo desarrollado a partir de normalización z-score. Error cuadrático medio de predicción de 0,58*

<b>Muestra</b>	<b>% Grasa experimental</b>	<b>%Grasa predicho</b>
186	53,74	53,65
189	53,76	54,21
194	51,56	51,75
195	54,65	54,49
52	47,98	47,88
53	47,6	47,93
55	51,16	51,09
56	49,11	48,47
82	54,97	54,87
86	53,95	53,65
89	54,25	53,76
92	54,06	54,62
751	55,45	54,91

a)



b)

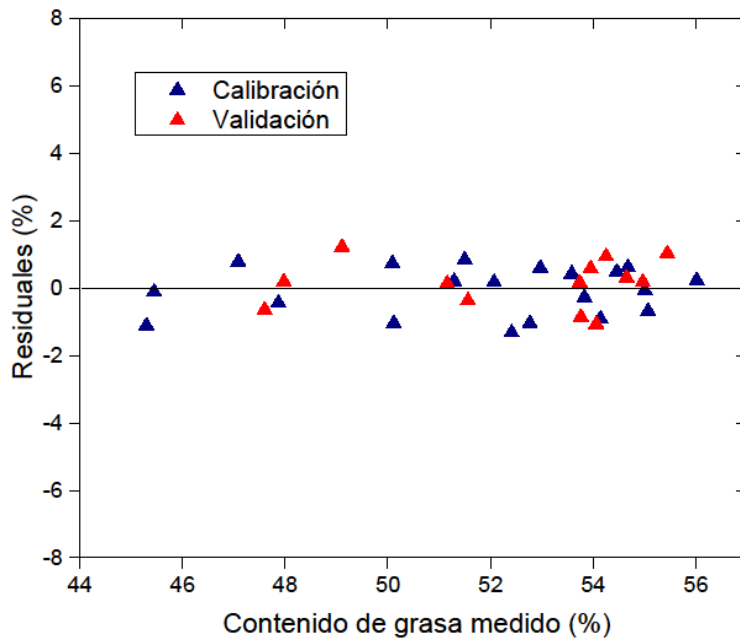


Figura 18. Gráfica de los residuales frente a los valores reales de grasa de cacao para los modelos bajo: a) Normalización por área, b) Normalización z-scores

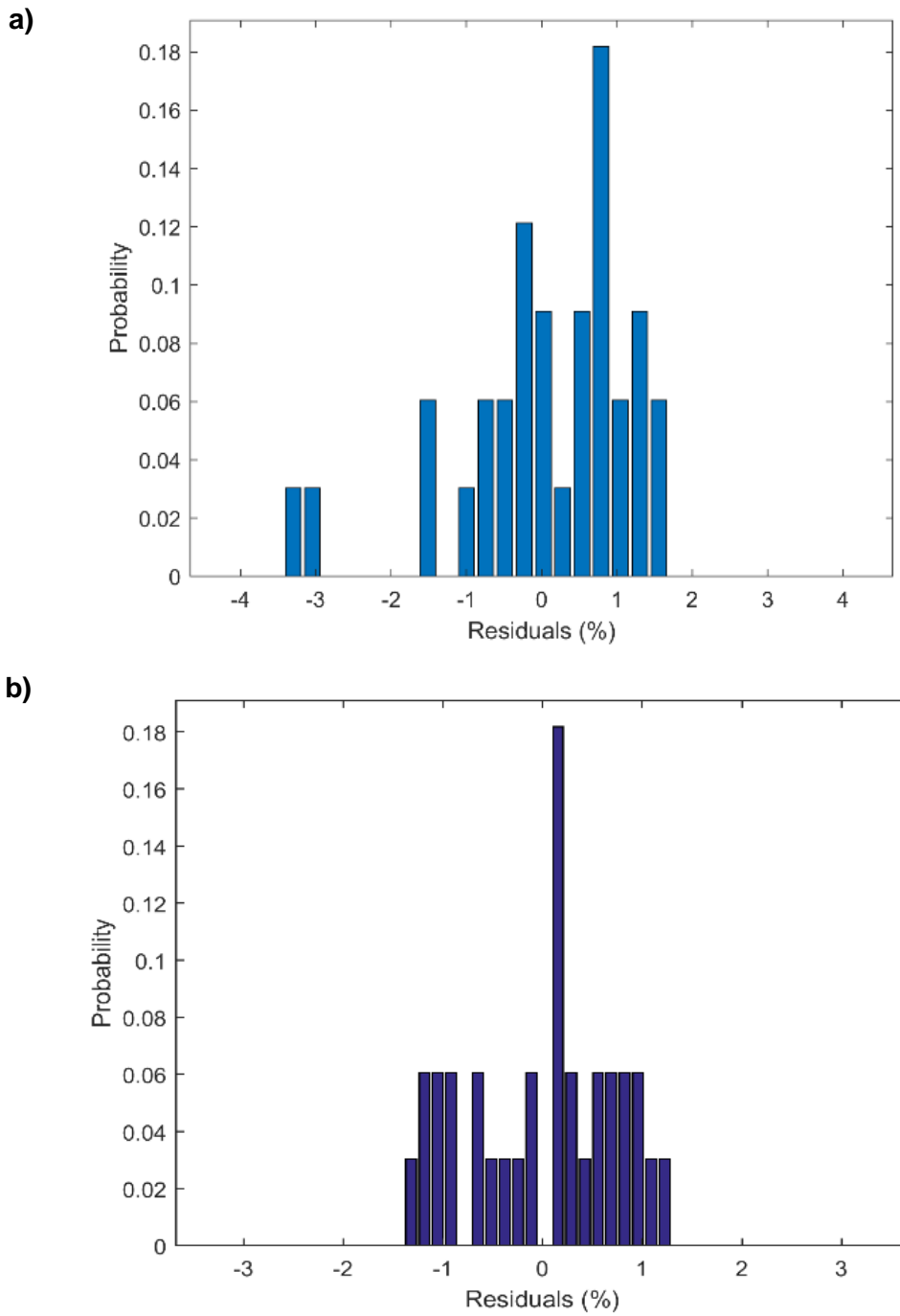


Figura 19. Histogramas de los residuales para los modelos bajo: a) Normalización por área, b) Normalización z-scores

## 6. Conclusiones

En este trabajo se han propuesto dos modelos predictivos para el contenido de grasa en cacaos colombianos, desarrollados por regresión de mínimos cuadrados parciales (PLS) a partir de 33 muestras de cacao de diferentes regiones del país, demostrando que es posible predecir esta propiedad de una forma fácil y rápida utilizando espectroscopia infrarroja (FTIR-ATR). El modelo construido con un pretratamiento de normalización SVN o z-score se obtuvo con errores cuadráticos medios de calibración y validación óptimos y un mejor coeficiente de regresión que el modelo generado a partir de normalización por área, por lo que se considera que esta es una metodología que podría usarse como alternativa a los métodos estándar, con la ventaja de que se reduciría el tiempo y costo implementados en el análisis.

Mediante el análisis de componentes principales (PCA), utilizando ocho componentes principales que explicaron un 93.62% de la variabilidad de los datos, fue posible realizar una aproximación en agrupaciones de las muestras de acuerdo con su lugar de origen, así como también según su porcentaje de grasa.

La viabilidad de implementación de esta metodología en la industria requeriría de la construcción de modelos más estables y robustos utilizando un mayor número de muestras, con el fin de que la estadística tenga más utilidad.

## 7. Divulgación Científica

Resumen del póster presentado en el XVII CONGRESO COLOMBIANO DE QUÍMICA

### **PREDICCIÓN DEL CONTENIDO DE GRASA EN CACAO A PARTIR DE ESPECTROSCOPIA INFRARROJA (FTIR-ATR) Y MÉTODOS QUIMIOMÉTRICOS**

Jennifer Steffens Ariza Ocaño, Karen Johanna Osorio Niño, Luis Javier López, Enrique Mejía-Ospino

steffan\_ariza@hotmail.com, karenosorio\_15@hotmail.com

Laboratorio de Espectroscopia Atómica y Molecular, Escuela de Química, Universidad Industrial de Santander. Bucaramanga, Colombia.

[emejia@uis.edu.co](mailto:emejia@uis.edu.co)

### Referencias Bibliográficas

- Aldana, R. (Julio-Agosto de 2017). Cacao de talla mundial. Colombia Cacaotera . *Periódico de la Federación Nacional de Cacaoteros No. 37.* , pág. 2.
- Cambrai, A., Marchioni, E., Julien-David, D., & Marcic, C. (2017). Discrimination of Cocoa Bean Origin by Chocolate Polyphenol Chromatographic Analysis and Chemometrics. *Food Analytical Methods, 10(6)*, 1991-2000.
- Cambrai, A., Marcic, C., Morville, S., Sae Houer, P., Bindler, F., & Marchioni, E. (2010). Differentiation of chocolates according to the cocoa's geographical origin using chemometrics. *Journal of agricultural and food chemistry, 58(3)*, 1478-1483.
- Cases, M. V., & Hens, A. G. (1988). *Técnicas analíticas de separación*. Reverté.
- Coitinho, T. B., Cassoli, L. D., Cerqueira, P. H., Da Silva, H. K., Coitinho, J. B., & Machado, P. F. (2017). Adulteration identification in raw milk using Fourier transform infrared spectroscopy. . *Journal of Food Science and Technology, 1-9*.
- Cozzolino, D., Parker, M., Damberg, R. G., Herderich, M., & Gishen, M. (2006). Chemometrics and Visible-Near Infrared spectroscopic monitoring of red wine fermentation in a pilot scale. *Biotechnology and bioengineering, 95(6)*, 1101-1107.
- Craig, A. P., Franca, A. S., & Oliveira, L. S. (2012). Evaluation of the potential of FTIR and chemometrics for separation between defective and non-defective coffees. *Food chemistry, 132(3)*, 1368-1374.
- Di Khanh, N. (2015). Advances in the extraction of anthocyanin from vegetables. *Journal of Food and Nutrition Sciences, 3(1-2)*, 126-134
- Dixit, V., Tewari, J. C., Cho, B. K., & Irudayaraj, J. M. (2005). Identification and quantification of industrial grade glycerol adulteration in red wine with Fourier transform infrared spectroscopy using chemometrics and artificial neural networks. *Applied spectroscopy, 59(12)*, 1553-1561.

- Eide, I., Zahlse, K., Kummernes, H., & Neverdal, G. (2006). Identification and quantification of surfactants in oil using the novel method for chemical fingerprinting based on electrospray mass spectrometry and chemometrics. *Energy & fuels*, 20(3), 1161-1164.
- Ellis, D. I., Broadhurst, D., & Goodacre, R. (2004). Rapid and quantitative detection of the microbial spoilage of beef by Fourier transform infrared spectroscopy and machine learning. *Analytica Chimica Acta*, 514(2), 193-201.
- Fedecacao. (2017). *Colombia alcanzó nuevo récord en producción de cacao*. Obtenido de [www.fedecacao.com.co](http://www.fedecacao.com.co)
- Fedecacao. (S.f.). *Récord histórico alcanza producción de cacao en 2016*. Obtenido de [www.fedecacao.com.co](http://www.fedecacao.com.co)
- Fernandez Pierna, J. A., Volery, P., Besson, R., Baeten, V., & Dardenne, P. (2005). Classification of modified starches by Fourier transform infrared spectroscopy using support vector machines. *Journal of agricultural and food chemistry*, 53(17), 6581-6585.
- Fernández, C. M. (2011). Quimiometría. *Universitat de València*, 15-18.
- Ferré, J. (2006). Calibración multivariante en análisis cuantitativo. *El modelo inverso. Técnicas de laboratorio*, (310), , 262-266.
- Geladi, P., & Kowalski, B. (1986). Partial least squares regression—a tutorial. *Anal. Chim. Acta*; 188, 19–32.
- Goodacre, R., & Anklam, E. (2001). Fourier transform infrared spectroscopy and chemometrics as a tool for the rapid detection of other vegetable fats mixed in cocoa butter. *Journal of the American Oil Chemists' Society*, 78(10), 993-1000.
- Holland, S. M. (2008). principal components analysis (PCA). *Department of Geology, University of Georgia, Athens, GA*, 30602-2501.
- Jahurul, M. H., Norulaini, N. A., Sahena, F., Jinap, S., Azmir, J., & Omar, A. M. (2013). Cocoa butter fats and possibilities of substitution in food products concerning cocoa varieties,

- alternative sources, extraction methods, composition, and characteristics. *Journal of Food Engineering*, 117(4), 467-476.
- Jee, M. (2002). *Oils and fats authentication*. CRC Press.
- Kalustian, P. (1985). Pharmaceutical and cosmetic uses of palm and lauric products. *Journal of the American Oil Chemists' Society*, 62(2), 431-433.
- Kowalski, B., Brown, S., & Vandeginste, B. (1987). *J. Chemometr.*
- Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H. (2015). Fast and neat-Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food chemistry*, 181, 152-159.
- Kumar, N., Bansal, A., Sarma, G. S., & Rawal, R. K. (2014). Chemometrics tools used in analytical chemistry: An overview. *Talanta*, 123, 186-199.
- Lache-García, A., Meléndez-Correa, L. V., & Orrego. (2010). *Predicción del Análisis SARA de Crudos Colombianos por Métodos Quimiométricos Utilizando Espectroscopia Infrarroja-ATR*. Bucaramanga : (Tesis de pregrado) Universidad Industrial de Santander.
- Macho, S., & Larrechi, M. S. (2002). Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *TrAC Trends in Analytical Chemistry*, 21(12), 799-806.
- Mackenzie, A. (2005). *Mathematics and statistics for life scientists*. Garland Science, 40.
- Maurer, N. E., & Rodriguez-Saona, L. (2013). Rapid assessment of quality parameters in cocoa butter using ATR-MIR spectroscopy and multivariate analysis. *Journal of the American Oil Chemists' Society*, 90(4), 475-481.
- McMullin, D., Mizaikoff, B., & Krska, R. (2015). Advancements in IR spectroscopic approaches for the determination of fungal derived contaminations in food crops. *Analytical and bioanalytical chemistry*, 407(3), 653-660.

- Miller, J. N., & Miller, J. C. (2002). *Estadística y quimiometría para química analítica*. 4a ed. Pearson Educación.
- Miller, K. B., Hurst, W. J., Payne, M. J., Stuart, D. A., Apgar, J., Sweigart, D. S., & Ou, B. (2008). Impact of alkalization on the antioxidant and flavanol content of commercial cocoa powders. *Journal of agricultural and food chemistry*, 56(18), 8527-8533.
- Naik, B., & Kumar, V. (2014). Cocoa butter and its alternatives: A review. *J. Biores. Eng. Technol*, 1. 7-17.
- Oliveros, D. (2017). *Measurement the Competitiveness of Cocoa Farmers in Santander*. Colombia.
- Pasto, D. J., & Johnson, C. R. (1981). *Determinación de estructuras orgánicas*. Reverté.
- Press, O. U. (2003). *Diccionario de química*. Complutense.
- Rios, R., Pessanha, M., Almeida, P., Viana, C., & Lannes, S. (2014). Aplicación de grasas en algunos productos alimenticios. . *Ciencia y Tecnología de Alimentos (Campinas)* , 34(1), 3-15.
- Rivero, E. G., Marder, L., Ferrão, M. F., Hermes, C. L., & Costa, A. (2008). Applied clean technologies to the study of pharmaceutical formulations. *In XIV Internatinal Conference on Industrial Engineering and Operations Management*. Rio de Janeiro.
- Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., & Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of pharmaceutical and biomedical analysis*, 44(3), 683-700.
- Romero, M. A. (s.f.). Desarrollo de nuevas tecnologías analíticas en el control de calidad de la industria farmacéutica. *Analyst*, 126, 2212-2217.
- Scafi, S. H., & Pasquini, C. (2001). Identification of counterfeit drugs using near-infrared spectroscopy. . *Analyst*, 126(12), 2218-2224.

- Skoog, D. A., Holler, F. J., & Timothy, A. D. (2001). *Principios de análisis instrumental* (No. 543.4/ 5). McGraw-Hill Interamericana de España.
- Teye, E., Huang, X. Y., Lei, W., & Dai, H. (2014). Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans. *Food research international*, 55, 288-293.
- Teye, E., Huang, X., Sam-Amoah, L. K., Takrama, J., Boison, D., Botchway, F., & Kumi, F. (2015). Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food chemistry*, 176, 403-410.
- Van Ruth, S. M., Rozijn, M., Koot, A., Garcia, R. P., Van der Kamp, H., & Codony, R. (2010). Authentication of feeding fats: classification of animal fats, fish oils and recycled cooking oils. *Animal Feed Science and Technology*, 155(1), 65-73.
- Vasconcelos, H., Saraiva, C., & De Almeida, J. M. (2014). Evaluation of the spoilage of raw chicken breast fillets using Fourier transform infrared spectroscopy in tandem with chemometrics. *Food and bioprocess technology*, 7(8), 2330-2341.
- Vázquez-Ovando, A., Molina-Freaner, F., Nuñez-Farfán, J., Betancur-Ancona, D., & Salvador-Figueroa, M. (2015). Classification of cacao beans (*Theobroma cacao* L.) of southern Mexico based on chemometric analysis with multivariate approach. *European Food Research and Technology*, 240(6), 1117-1128.
- Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica chimica acta*, 601(1), 77-86.
- Wold, S., & Sjöström, M. (1998). Chemometrics, present and future success. *Chemometrics and Intelligent Laboratory Systems*, 44(1-2), 3-14.
- Wu, Z., Li, H., & Tu, D. (2015). Application of Fourier Transform Infrared (FT-IR) Spectroscopy combined with chemometrics for analysis of rapeseed oil adulterated with refining and purificating waste cooking oil. *Food analytical methods*, 8(10), 2581-2587.
- Xu, Q. S., & Liang, Y. Z. (2001). Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, 56(1), 1-11.

Zuluaga, C. M., Moreno, C. D., & Quicazán, M. C. (2011). Quimiometría aplicada a la diferenciación por origen de productos de las abejas. . *Alimentos Hoy*, 20(24), 23-46.