



UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA
CARRERA DE BIOTECNOLOGÍA

DESARROLLO DE UN MODELO MATEMÁTICO MULTIVARIADO PREDICTIVO
PARA LA CUANTIFICACIÓN SIMULTÁNEA DE GLUCOSA, FRUCTOSA Y
SACAROSA EN BEBIDAS AZUCARADAS MEDIANTE ESPECTROSCOPÍA
INFRARROJA POR TRANSFORMADA DE FOURIER (FTIR)

Trabajo de titulación previo a la obtención del
título de Ingeniera Biotecnóloga

AUTORA: ALLISON MARÍA SOL GRANDA ROMERO

TUTOR: DR. PABLO WILSON ARÉVALO MOSCOSO, PhD.

Cuenca - Ecuador

2026

**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, Allison María Sol Granda Romero con documento de identificación N° 0706178787
manifiesto que:

Soy la autora y responsable del presente trabajo; y, autorizo a que sin fines de lucro la
Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera
total o parcial el presente trabajo de titulación.

Cuenca, 9 de febrero del 2026

Atentamente,



Allison María Sol Granda Romero

0706178787

**CERTIFICADO CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo, Allison María Sol Granda Romero con documento de identificación N° 0706178787, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autora del Trabajo experimental: “Desarrollo de un modelo matemático multivariado predictivo para la cuantificación simultánea de glucosa, fructosa y sacarosa en bebidas azucaradas mediante espectroscopía infrarroja por transformada de Fourier (FTIR)”, el cual ha sido desarrollado para optar por el título de: Ingeniera Biotecnóloga, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 9 de febrero del 2026

Atentamente,



Allison María Sol Granda Romero

0706178787

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Pablo Wilson Arévalo Moscoso con documento de identificación N° 0102156957, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: DESARROLLO DE UN MODELO MATEMÁTICO MULTIVARIADO PREDICTIVO PARA LA CUANTIFICACIÓN SIMULTÁNEA DE GLUCOSA, FRUCTOSA Y SACAROSA EN BEBIDAS AZUCARADAS MEDIANTE ESPECTROSCOPIA INFRARROJA POR TRANSFORMADA DE FOURIER (FTIR), realizado por Allison María Sol Granda Romero con documento de identificación N° 0706178787 obteniendo como resultado final el trabajo de titulación bajo la opción de Trabajo experimental que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 9 de febrero del 2026

Atentamente,



Dr. Pablo Wilson Arévalo Moscoso, PhD.

0102156957

DEDICATORIA

El presente trabajo de titulación está dedicado a quienes fueron mi rayo de luz cuando sentía que todo se volvía oscuridad.

Dedicado a mis padres, José Luis y Fabiola, que desde el inicio de mi trayectoria académica me apoyaron incondicionalmente. Se aseguraron de que nunca me falte una llamada, un mensaje, o un abrazo alentador.

Dedicado a mi hermano mayor, Carlos, quien siempre fue paciente conmigo al explicarme aquello que no entendía una y otra vez, y quien todos los días preguntaba por mí.

Dedicado a mi abuelita, Rosa, quien me extrañaba todos los días y cada vez que regresaba a mi hogar me recibía con dulzura.

Dedicado a mis gatos: Pelusa, Mimi, Nino y Mochi; quienes me enseñaron que no existe una conexión y amor más puro.

AGRADECIMIENTOS

Primero, agradezco a Dios y a la Virgen María, por ser mi soporte en todo momento.

A mis padres, José Luis y Fabiola, por su infinito amor, por todos los sacrificios que han hecho para permitirme llegar hasta aquí y por motivarme a luchar por mis sueños.

A mi hermano Carlitos, por aconsejarme cuando lo necesitaba y darme la mano en todo momento.

A mi abuelita, por su calidez, por su sabiduría y por creer en mí.

A mis gatos, por brindarme alegría desde que llegaron a mi vida.

A Brandon, por su amistad sincera; por todas las risas y aventuras que compartimos; por todo su apoyo y por hacer de esta etapa una experiencia inolvidable.

A mis amigas y amigos, quienes han celebrado mis triunfos como si fuesen suyos y han estado en las buenas y en las malas.

A mi tutor, el Dr. Pablo Arévalo, por compartir sus conocimientos y haber sido mi guía durante el desarrollo de este trabajo.

Finalmente, agradezco a mi persona por soñar en grande, por encontrar fortaleza cuando sentía que desvanecía y por haberme amado más que nunca.

TABLA DE CONTENIDOS

RESUMEN	11
ABSTRACT	13
INTRODUCCIÓN.....	15
1.1. Planteamiento del problema.....	16
1.2. Pregunta de Investigación.....	17
1.3. Justificación	17
1.4. Limitaciones.....	19
1.5. Objetivos.....	19
1.5.1. Objetivo General	19
1.5.2. Objetivos Específicos.....	19
1.6. Hipótesis	20
1.7. Variables.....	20
1.7.1. Independientes.....	20
1.7.2. Dependiente.....	20
1.7.3. Intervinientes.....	21
1.7.4. Extrañas.....	21
MARCO TEÓRICO	22
2.1. Antecedentes de la Investigación.....	22
2.2. Bases Teóricas.....	23
2.2.1. Espectroscopía Infrarroja por Transformada de Fourier (FTIR).....	23
2.2.2. Región diagnóstica y región de las huellas digitales.....	24
2.2.3. Señal residual del <i>background</i>	26
2.2.4. Diseño experimental de mezclas	26
2.2.5. Preprocesamiento de datos espectrales	27
2.2.6. Modelamiento matemático y quimiometría	29
2.3. Marco Conceptual.....	31

2.3.1. Espectroscopía Infrarroja por Transformada de Fourier (FTIR).....	31
2.3.2. Glucosa.....	31
2.3.3. Fructosa.....	32
2.3.4. Sacarosa.....	32
2.3.5. Modelo matemático multivariado.....	32
2.3.6. Regresión por mínimos cuadrados parciales.....	32
2.3.7. Preprocesamiento espectral.....	33
2.3.8. Validación del modelo.....	33
MARCO METODOLÓGICO.....	34
3.1. Nivel de Investigación.....	34
3.2. Diseño de Investigación.....	34
3.3. Diseño de Experimento.....	34
3.4. Técnicas e Instrumentos de Recolección de Datos.....	36
3.5. Técnicas de Procesamiento y Análisis de Datos.....	36
3.6. Protocolo a implementar.....	37
3.6.2. Corrida de las muestras en el equipo FTIR-ATR.....	38
3.6.3. Métodos estadísticos.....	44
ANÁLISIS DE RESULTADOS Y DISCUSIÓN.....	46
4.1. Resultados experimentales obtenidos.....	46
4.1.1. Análisis de componentes principales.....	46
4.1.2. Selección de rangos óptimos.....	49
4.1.3. Modelo de regresión por mínimos cuadrados parciales (PLS).....	50
4.1.4. Error cuadrático medio (RMSE) y coeficiente de determinación (R^2).....	53
4.1.5. Validación del modelo con muestras de bebidas comerciales.....	55
4.2. Discusión.....	57
CONCLUSIONES Y RECOMENDACIONES.....	60
5.1. Conclusiones.....	60

5.2. Recomendaciones	60
REFERENCIAS	62
ANEXOS	71

ÍNDICE DE TABLAS

Tabla 1.....	20
Tabla 2.....	21
Tabla 3.....	35
Tabla 4.....	51
Tabla 5.....	52
Tabla 6.....	53
Tabla 7.....	55
Tabla 8.....	56

ÍNDICE DE FIGURAS

Figura 1.....	24
Figura 2.....	25
Figura 3.....	25
Figura 4.....	31
Figura 5.....	37
Figura 6.....	39
Figura 7.....	39
Figura 8.....	41
Figura 9.....	41
Figura 10.....	43
Figura 11.....	43
Figura 12.....	46
Figura 13.....	47
Figura 14.....	48
Figura 15.....	51
Figura 16.....	53
Figura 17.....	54
Figura 18.....	54
Figura 19.....	56

RESUMEN

La cuantificación de azúcares en bebidas es un proceso importante dentro de la industria alimentaria, especialmente para el control de calidad y el cumplimiento de la normativa. Los métodos tradicionales como la cromatografía suelen ser costosos y lentos, por lo que este trabajo propone el desarrollo de un modelo matemático multivariado predictivo basado en espectroscopía infrarroja por transformada de Fourier (FTIR-ATR) destinado a la cuantificación simultánea de glucosa, fructosa y sacarosa en bebidas azucaradas, como una alternativa rápida y más accesible.

El proceso conllevó un diseño de 32 mezclas estándar con concentraciones variables de los tres azúcares, basándose en un diseño experimental de mezclas con restricciones. Luego, los espectros se obtuvieron mediante el análisis en el equipo FTIR-ATR y se preprocesaron aplicando corrección de línea base por ALS, normalización, suavizado por Savitzky–Golay y derivadas (primera y segunda). Se seleccionó la región espectral de la huella digital de los carbohidratos, comprendida entre de $850\text{--}1200\text{ cm}^{-1}$, para construir el modelo, ya que esta región reducía el ruido y mejoraba la selectividad. Posteriormente, el análisis de componentes principales (PCA) confirmó la estructuración de los datos, y se continuó con la construcción del modelo de regresión por mínimos cuadrados parciales (PLS) con 12 componentes.

Los resultados evidenciaron mostraron un error cuadrático medio (RMSE) entre $0.74\text{--}0.81$ y un coeficiente de determinación (R^2) promedio de 0.80 para los tres azúcares en la fase de prueba, indicando una capacidad predictiva moderada. No obstante, la validación con bebidas comerciales demostró limitaciones claras del modelo al extrapolar concentraciones fuera del rango de calibración, obteniendo predicciones erróneas que incluían valores negativos. Dado esto, a pesar de que el enfoque sea prometedor, es necesario ampliar el conjunto de datos de calibración con mayor variabilidad en las

concentraciones y optimizar el modelo para mejorar su robustez y aplicabilidad en matrices complejas.

Palabras clave: Modelo multivariado predictivo, FTIR-ATR, cuantificación, glucosa, fructosa, sacarosa.

ABSTRACT

The quantification of sugars in beverages is an important process within the food industry, especially for quality control and regulatory compliance. Traditional methods such as chromatography are often costly and time-consuming, which is why this work proposes the development of a predictive multivariate mathematical model based on Fourier transform infrared spectroscopy (FTIR-ATR) for the simultaneous quantification of glucose, fructose, and sucrose in sugar-sweetened beverages, as a faster and more accessible alternative.

The process involved the design of 32 standard mixtures with varying concentrations of the three sugars, based on a constrained mixture experimental design. The spectra were then obtained by analysis using the FTIR-ATR equipment and were preprocessed by applying baseline correction (ALS), normalization, smoothing (Savitzky–Golay), and derivative transformations (first and second). The spectral region of the carbohydrate fingerprint, ranging from 850–1200 cm^{-1} , was selected for building the model, as this region reduced noise and improved selectivity. Subsequently, principal component analysis (PCA) confirmed the structuring of the data, and the construction of a partial least squares (PLS) regression model with 12 components was carried out.

The results showed a root mean square error (RMSE) between 0.74–0.81 and an average coefficient of determination (R^2) of 0.80 for the three sugars in the testing phase, indicating a moderate predictive capability. However, validation with commercial beverages demonstrated clear limitations of the model when extrapolating concentrations outside the calibration range, leading to erroneous predictions that included negative values. Therefore, although the approach is promising, it is necessary to expand the calibration dataset with greater concentration variability and optimize the model to improve its robustness and applicability in complex matrices.

Keywords: Predictive multivariate model, FTIR-ATR, quantification, glucose, fructose, sucrose.

CAPÍTULO I

INTRODUCCIÓN

En la industria alimentaria surge la necesidad de métodos rápidos y confiables que cuantifiquen azúcares, componentes que influyen en la calidad nutricional y comercial de diversos alimentos y bebidas. Específicamente, la glucosa, fructosa y sacarosa son los azúcares presentes en mayor proporción en bebidas azucaradas (Ando et al., 2023), por lo que su determinación es esencial para el control de calidad y el cumplimiento de normativas.

Actualmente, se buscan herramientas de análisis que complementen o reemplacen métodos tradicionales como la cromatografía o titulación, por su mayor inversión de tiempo, reactivos y equipos. La espectroscopía infrarroja por transformada de Fourier (FTIR) es una de las técnicas analíticas más atractivas para el estudio de los componentes mayoritarios de un producto (Mondragón, 2020). En combinación con modelos matemáticos multivariados, son la opción que más se alinea con una analítica verde y automatización de procesos. Asimismo, permite detectar adulteraciones en los productos, asegurando transparencia y velando por la seguridad de los consumidores.

El propósito de la presente investigación es desarrollar un modelo matemático multivariado predictivo que permita cuantificar simultáneamente glucosa, fructosa y sacarosa a partir del espectro proporcionado por la técnica de espectroscopía infrarroja, identificando rangos espectrales óptimos que orienten a la construcción de modelos de regresión y la validación de su capacidad predictiva.

Por otro lado, el enfoque teórico se centra en la quimiometría aplicada al análisis espectral, especialmente a los principios de modelamiento multivariado y la interpretación de correlaciones entre señales espectrales y concentraciones. La investigación se cimenta en la naturaleza estructural de los carbohidratos estudiados, sus

características espectrales y los fundamentos de la espectroscopía infrarroja por transformada de Fourier. De esta forma, se establecen los criterios necesarios para la interpretación de los resultados.

1.1. Planteamiento del problema

En la industria alimentaria, caracterizar químicamente los productos resulta imperante para garantizar la calidad y cumplir la normativa de inocuidad establecida por entidades regulatorias. El contenido de azúcares se encuentra entre los parámetros que requieren un control riguroso en bebidas naturales y procesadas, donde la glucosa, fructosa y sacarosa son los principales ingredientes. Según la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), el consumo global de azúcares ha aumentado considerablemente en las últimas décadas, ignorando las recomendaciones de la Organización Mundial de la Salud, que establecen un máximo del 10 % de las calorías diarias de azúcares libres (World Health Organization, 2015). Además, se prevé que, en los próximos diez años, el consumo mundial de azúcar crezca un 1.2 % anual y alcance los 198 millones de toneladas en 2033 (OECD-FAO, 2009).

Esta problemática se acentúa en bebidas procesadas, datos del World Bank (2020) indican que las bebidas azucaradas representan cerca del 28% del consumo mundial de azúcares. En Ecuador, Mayorga et al. (2022) reportan que, entre los años 2014 – 2019, las bebidas carbonatadas presentaron el 83% del consumo total de bebidas azucaradas, y los jugos el 10%. Aquello se relaciona directamente con el hecho de que, en Ecuador, la cifra de infantes y adolescentes que padecen de sobrepeso y obesidad asciende al 28% (UNICEF Ecuador, 2024).

La cuantificación de glucosa, fructosa y sacarosa se lleva a cabo a través de cromatografía iónica o cromatografía líquida de alta resolución (HPLC) (Pérez-López, 2023). A pesar de que las técnicas mencionadas anteriormente son precisas, requieren

tiempos prolongados de análisis, emplean reactivos costosos, generan residuos y se necesita personal capacitado (Prada, 2015). Asimismo, puede haber cambios en la posición o anchura del pico por las variabilidades de la muestra (Worley & Powers, 2012). En consecuencia, existen brechas en la vigilancia y un incremento en el riesgo de que la composición del producto presente inconsistencias.

Frente a esta problemática, la espectroscopía infrarroja por transformada de Fourier (FTIR) es una solución prometedora. No obstante, la cuantificación simultánea de glucosa, fructosa y sacarosa precisa de modelos matemáticos multivariados que procesen el solapamiento natural de señales espectrales de los carbohidratos. La ausencia de modelos predictivos validados para cuantificar azúcares es una barrera para implementar la espectroscopía FTIR en la industria alimentaria, ya que, sin un modelo adecuado, la técnica no podría utilizarse para el control de calidad, verificación de etiquetado o la detección de adulteraciones.

1.2. Pregunta de Investigación

¿Existen relaciones funcionales entre las longitudes de onda de los espectros FTIR y las concentraciones de glucosa, fructosa y sacarosa que permitan generar modelos de regresión multivariada para la predicción simultánea de estos azúcares en muestras de bebidas?

1.3. Justificación

La industria alimentaria y los laboratorios dependen de técnicas cromatográficas para la cuantificación de azúcares en productos comerciales, las cuales no son viables por sus costos elevados e impacto ambiental. En este contexto, desarrollar métodos más rápidos y accesibles se ha convertido en una necesidad, por lo que el propósito de esta investigación es proponer un modelo matemático multivariado predictivo a partir de

espectros FTIR que cuantifique de manera simultánea glucosa, fructosa y sacarosa de forma eficiente y reproducible en diversos entornos.

El estudio es relevante en el ámbito científico debido a que fusiona la espectroscopía infrarroja por transformada de Fourier (FTIR) con quimiometría avanzada, una rama en expansión, a fin de maximizar la productividad de la información espectral al tratar datos complejos que pueden ser correlacionados con características de la muestra (Chavarro & Vargas, 2022), mediante algoritmos de regresión y selección de variables. A nivel contemporáneo, hay una creciente demanda de análisis rápidos y precisos de bebidas azucaradas por el aumento de su comercialización, donde se busca detectar adulteraciones y monitorear cambios químicos durante la producción (Guerrero Niño et al., 2025).

En cuanto al aporte teórico, el estudio fortalecerá la comprensión de la relación entre las señales espectrales y los grupos funcionales de carbohidratos en matrices líquidas, esclareciendo los factores que influyen en el modelamiento matemático. Desde el punto de vista práctico, proporcionará una herramienta que puede ser implementada en laboratorios, aminorando los tiempos operativos. Del mismo modo, pavimenta el camino para el desarrollo de métodos automatizados que faciliten la transición a estudios a gran escala.

Los beneficiarios directos serán laboratorios de control de calidad, centros de investigación y la industria alimentaria, quienes podrán optimizar sus procesos y brindar resultados fiables. Los beneficiarios indirectos serán los usuarios, ya que se garantizará que las bebidas que adquieran sean auténticas nutricionalmente y se promoverá un consumo responsable.

1.4. Limitaciones

Durante el desarrollo del presente estudio se prevé la presencia de algunas limitaciones que podrían influenciar la ejecución del experimento y/o resultados de este. En primer lugar, la alta afluencia de estudiantes e investigadores en los laboratorios institucionales puede restringir la disponibilidad de espacios, equipos e instrumentos en los horarios requeridos. Esta situación puede generar demoras en la realización y seguimiento de los análisis.

Por otro lado, la detección de los analitos objetivo puede verse afectada por otros componentes de las bebidas, lo cual puede generar interferencias espectrales, alterar la selectividad del modelo y afectar el error de predicción. En tal caso, la precisión del modelo disminuye, al igual que la veracidad de los resultados.

1.5. Objetivos

1.5.1. Objetivo General

Desarrollar un modelo matemático multivariado predictivo, basado en espectroscopía infrarroja por transformada de Fourier (FTIR), para la cuantificación simultánea de las concentraciones de glucosa, fructosa y sacarosa en bebidas azucaradas, como una técnica alternativa para el análisis de calidad.

1.5.2. Objetivos Específicos

- Determinar los rangos espectrales óptimos mediante el análisis de las bandas vibracionales características de glucosa, fructosa y sacarosa, a fin de procesarlos.
- Formular modelos matemáticos multivariados predictivos a partir de los datos espectrales obtenidos en el laboratorio, para realizar la cuantificación de los tres carbohidratos.

- Evaluar la precisión y la capacidad predictiva de los modelos construidos utilizando muestras de bebidas azucaradas, con propósito de validar la exactitud de la cuantificación de glucosa, fructosa y sacarosa.

1.6. Hipótesis

Existe una relación funcional entre las longitudes de onda y la concentración de los carbohidratos, permitiendo desarrollar modelos de predicción multidimensionales precisos.

1.7. Variables

1.7.1. Independientes

Múltiples variables independientes correspondientes a las longitudes de onda de los espectros.

Tabla 1

Tabla de operacionalización correspondiente a las variables independientes.

Variable	Dimensión	Indicador
Longitud de onda	Número de longitudes de onda seleccionadas	Unidad de medida de absorbancia, rango espectral de 400 a 4000 cm ⁻¹

Nota. Elaboración propia.

1.7.2. Dependiente

Variable tridimensional correspondiente a glucosa, fructosa y sacarosa.

Tabla 2

Tabla de operacionalización correspondiente a la variable dependiente.

Variable	Dimensión	Indicador
Concentración	$N * 2$, donde N es el número total de muestras.	Unidad de concentración (% v/v)

Nota. Elaboración propia.

1.7.3. Intervinientes

- Condiciones de preparación de la muestra
- Analista
- Solapamiento espectral entre glucosa, fructosa y sacarosa.

1.7.4. Extrañas

- Humedad del ambiental al realizar las mediciones.
- Temperatura durante el análisis.
- Pureza de los reactivos.
- Fallas en el equipo FTIR (mal funcionamiento o desajuste).

CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes de la Investigación

En los últimos años, la espectroscopía infrarroja por transformada de Fourier se ha posicionado como una de las técnicas más eficientes y confiables para el análisis de los ingredientes de alimentos y bebidas presentes en el mercado. En diversos estudios se ha evidenciado su utilidad, en combinación con métodos quimiométricos, para la identificación, cuantificación y autenticación de azúcares en matrices líquidas, como jugos y bebidas procesadas.

Entre los primeros estudios destaca la investigación de Duarte et al. (2002), en la cual evaluaron la aplicación de FTIR-ATR y regresión PLS para cuantificar glucosa, fructosa y sacarosa en jugo de mango a lo largo del proceso de maduración; describieron que modelos PLS aplicados a derivadas de los espectros FTIR permitieron cuantificar fructosa y sacarosa con errores de predicción bajos. Sin embargo, la cuantificación de glucosa tuvo mayor error cuando la concentración era muy baja en etapas tardías de maduración. El estudio demostró que la combinación de FTIR con PLS es viable como un método rápido y no destructivo para el análisis de azúcares en jugos.

También, Dhaulaniya et al. (2020) desarrollaron un modelo FTIR-quimiométrico que identificaba y cuantificaba sacarosa como adulterante en jugos de manzana; encontraron que la región espectral entre 1200–900 cm^{-1} es el “*fingerprint*” de los carbohidratos, ya que observaron picos asociados a azúcares simples, incluyendo uno distintivo a 997 cm^{-1} en las muestras adulteradas. Además, los modelos que construyeron mediante PLS y PCR mostraron coeficientes de determinación superiores a 0.99, confirmando que los enfoques multivariados arrojan estimaciones precisas.

De manera similar, Calle et al. (2022) evaluaron la técnica de FTIR y quimiometría para detectar adulteraciones de jugos de fruta con azúcares de bajo costo. Su estudio concluyó que modelos basados en PLS-R discriminan bebidas auténticas de adulteradas con alta precisión. Del mismo modo, enfatizaron la necesidad de métodos rápidos y no destructivos para el control de calidad en la industria alimentaria. Asimismo, al analizar la aplicación de FTIR y análisis multivariado en la detección de miel adulterada en mercados de Etiopía, Damto et al. (2023) acordaron que se puede clasificar el agente adulterante mediante un análisis de conglomerados a partir de diferencias espectrales, y que el FTIR, fusionado con análisis multivariante de los datos, es una valiosa huella digital para diferenciar la miel pura y adulterada.

Por otro lado, los modelos matemáticos predictivos basados en FTIR cuantifican de forma sólida diversos componentes en alimentos mediante calibraciones multivariadas. También, enfoques híbridos entre selección de variables y PLS mejoran la exactitud de los modelos, particularmente en matrices complejas, lo cual es aplicable a bebidas con múltiples azúcares presentes.

2.2. Bases Teóricas

2.2.1. Espectroscopía Infrarroja por Transformada de Fourier (FTIR)

Es una técnica analítica que usa un interferómetro para obtener un espectro mediante la transformada de Fourier (Skoog et al., 2022). Permite identificar grupos funcionales y estudia estructuras químicas al registrar la interacción entre la radiación infrarroja y la materia, de manera que cada banda en el espectro es asignada según la vibración de la estructura molecular (Stuart, 2021). Las ventajas de los espectrómetros FTIR, descritas por Christian et al. (2014), son: mayor capacidad, mayor relación de señal a ruido y medición simultánea en todas las longitudes de onda.

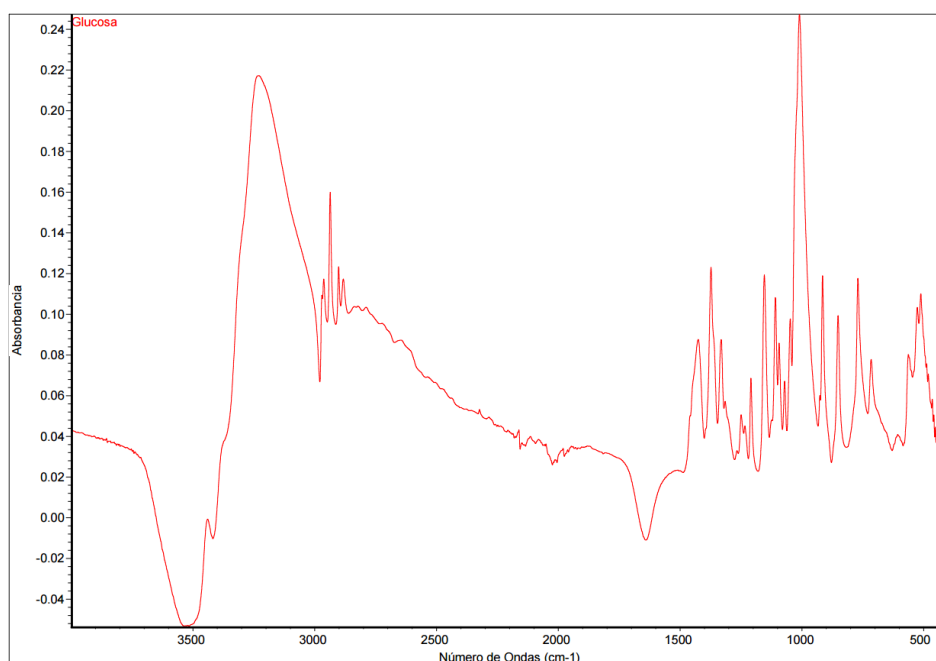
2.2.2. Región diagnóstica y región de las huellas digitales

La región diagnóstica se encuentra entre $4000\text{--}1300\text{ cm}^{-1}$ y contiene bandas de estiramiento de los grupos funcionales principales como O–H, C–H, C=O, N–H (Workman & Weyer, 2007). Por las señales simples que provee, sirve para identificar familias químicas, pero no para diferenciar dos o más analitos con grupos funcionales similares.

La región de la huella digital ($1300\text{--}400\text{ cm}^{-1}$) es un conjunto de bandas altamente específicas para cada molécula por las vibraciones combinadas, lo que permite diferenciar compuestos con estructuras similares (Stuart, 2021). Para el presente caso hay un especial interés en la región de los carbohidratos, los cuales muestran señales intensas entre 1200 y 900 cm^{-1} por las vibraciones C–O, C–C y C–O–H (Dhaulaniya et al., 2020; Szymanska-Chargot & Zdunek, 2013).

Figura 1

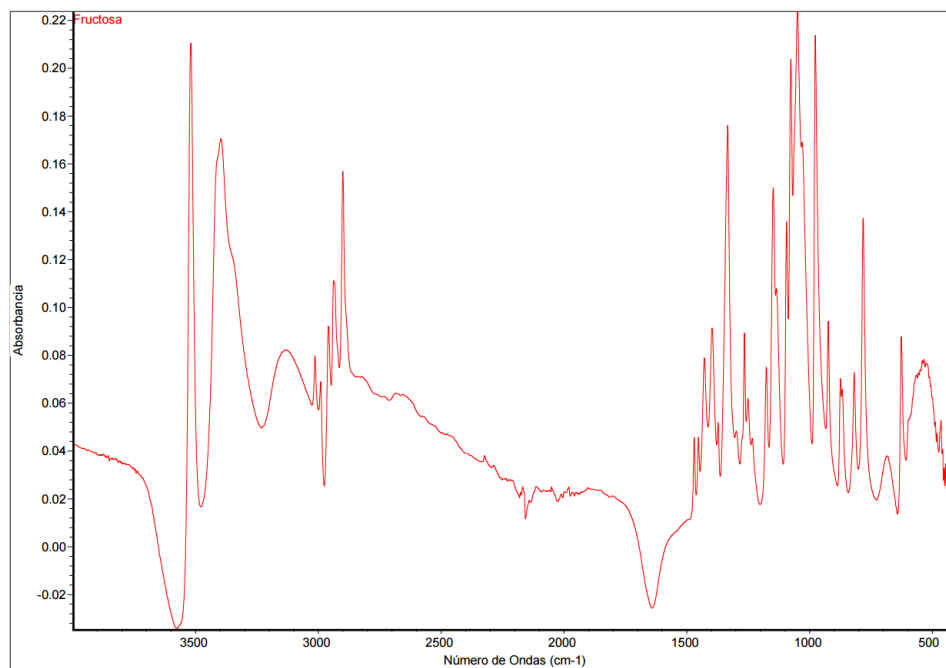
Espectro de glucosa



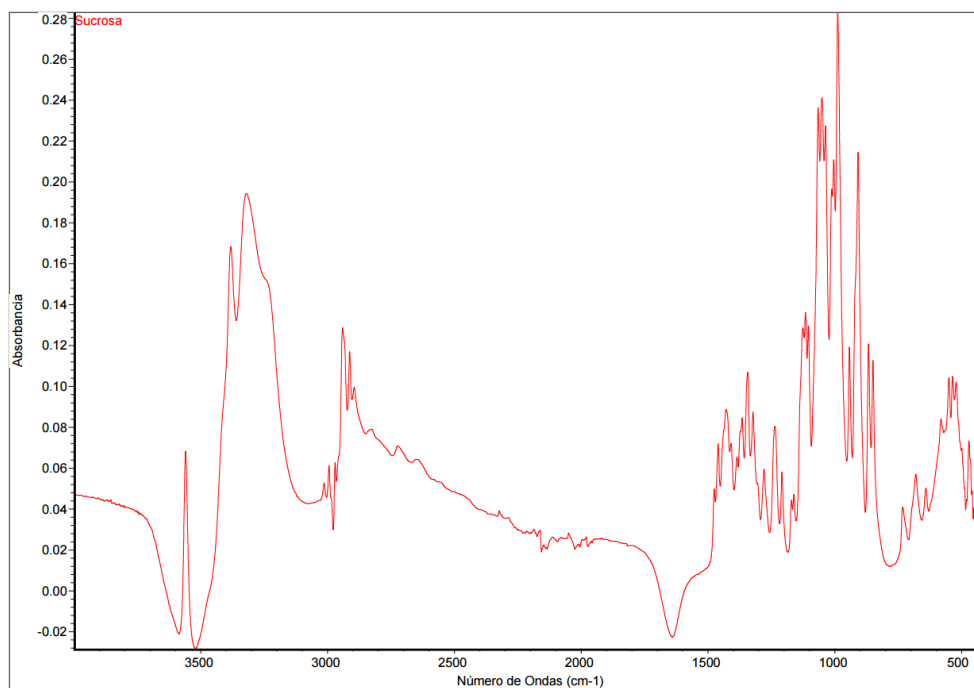
Nota. Propia autoría.

Figura 2

Espectro de fructosa

*Nota.* Propia autoría**Figura 3**

Espectro de sucrosa



Nota. Propia autoría

2.2.3. Señal residual del *background*

El *background* es el espectro de referencia o ambiente que incluye absorbancias de la atmósfera, instrumento y materiales ópticos. Si el fondo no coincide con las condiciones de la muestra, se generarán bandas no deseadas (Workman & Weyer, 2007). Por ende, antes de obtener el espectro de la muestra, se sustrae automáticamente el *background* para evitar interferencias.

2.2.4. Diseño experimental de mezclas

Es un diseño estadístico empleado para mejorar las características de una combinación de componentes, debido a que la respuesta del sistema depende a las proporciones de estos (Becerra & Goos, 2021). Los factores dependen uno del otro y las proporciones suman ya sea 100 % o 1 (Bolaños-Quirós et al., 2024). Es ideal al estudiar formulaciones o mezclas de compuestos (Cornell, 2011), como azúcares en soluciones acuosas.

En un diseño de mezclas estándar, la condición es la siguiente:

$$\sum_{i=1}^q x_i = 1$$

Donde:

X_i es la proporción de cada componente en la mezcla.

q es el número total de componentes.

No obstante, algunas ocasiones los componentes no pueden variar sin restricciones de 0 - 1 por ajustarse a algunas limitaciones. Dado esto, se sugiere que:

$$lc \leq x_i \leq uc$$

Aquí x_i representa la proporción de cada componente, lc es la mínima restricción del i -ésimo componente y uc es la proporción máxima del i -ésimo componente.

2.2.5. Preprocesamiento de datos espectrales

2.2.5.1 Corrección de línea base con ALS (*Asymmetric Least Squares*)

La línea base de un espectro son las regiones de este en donde no existen absorciones significativas (Moros, 2007). Se ve afectada por la calidad del *background*, calidad de la muestra, la preparación de la muestra, tipo de accesorio empleado, la estabilidad térmica del sistema, el alineamiento del divisor de haz y la purga del sistema (Moros, 2007).

La corrección de la línea base consiste en ajustar una curva suave a la señal para corregir desplazamientos de esta, de manera que se mantengan los picos de interés. En el presente caso se realizará por mínimos cuadrados ponderados, que penaliza desviaciones en una dirección para separar el fondo y el espectro (Baek et al., 2015).

2.2.5.2. Normalización de datos

La normalización permite equilibrar las características de los datos y reducir valores atípicos en el conjunto de datos originales, mejorando la calidad de los datos y, por ende, la capacidad de los modelos predictivos (Cabello-Solorzano et al., 2023). En la investigación se llevará a cabo una normalización por área, para que todos los espectros cuenten con la misma área total. Para esto, se aplicará la siguiente fórmula:

$$A_{\lambda, norm} = \frac{A_{\lambda}}{\sum A_{\lambda}}$$

Donde:

$A(\lambda)$ es la absorbancia en una longitud de onda específica.

$\int A(\lambda) d\lambda$ es la integral del espectro

$A(\lambda)/A(\lambda_m)$ es la absorbancia normalizada

2.2.5.3. Suavizado

El suavizado por filtros como Savitzky–Golay reduce el ruido sin comprometer la forma de las bandas vibracionales, de tal forma que se diferencien las señales de mejor manera (Krishnan & Seelamantula, 2013; Savitzky & Golay, 1964). La ecuación general es:

$$y_i^{(s)} = \sum_{j=-m}^m c_j y_{i+j}$$

Donde:

$y_i^{(s)}$ es el valor suavizado en la posición i .

y_{i+j} son los valores originales de los datos.

c_j son los coeficientes de convolución obtenidos del ajuste de polinomios de orden k .

m es el número de puntos vecinos usados en el ajuste.

2.2.5.4. Segunda derivada

La segunda derivada aplicada a espectroscopía resalta cambios en la pendiente del espectro, separa bandas superpuestas y, en el análisis de datos, detecta cambios bruscos en la señal (Savitzky & Golay, 1964). Se aplica la siguiente fórmula:

$$y_i^{(n)} = \sum_{j=-m}^m (c_j)^2 y_{i+j}$$

Donde:

$(\square\square)^2$ son los coeficientes de la segunda derivada.

$\square\square$ (") es la segunda derivada del filtro.

$\square\square + j$ son los valores originales de la señal.

2.2.6. Modelamiento matemático y quimiometría

2.2.6.1. Análisis de componentes principales (PCA)

Es un método estadístico de reducción de la dimensión (número de variables) a sus características más importantes, denominados componentes principales, perdiendo poca cantidad de información (Greenacre et al., 2022).. Se calculan dos componentes principales, el primero es la mejor dirección para proyectar los puntos y el segundo contendrá la máxima cantidad de la variabilidad restante. Así, Moros (2007) explica que su objetivo es "encontrar un nuevo conjunto de ejes ortogonales en el que la varianza de los datos sea máxima".

2.2.6.2. Regresión por mínimo cuadrados parciales (PLS) en análisis cuantitativo

Es una técnica del modelado multivariado recomendada cuando existe un gran número de variables predictoras altamente correlacionadas y un conjunto de variables de respuesta, ya que selecciona las variables y reduce la dimensionalidad (Lê Cao et al., 2011). El modelo se expresa como:

$$\hat{y} = b_0 + \sum_{i=1}^p b_i x_i$$

Donde:

p es el número de números de onda.

b_i es la matriz de coeficientes de regresión obtenida a partir de los componentes usados.

x_i son las absorbancias correspondientes a cada número de onda.

2.2.6.3. Error cuadrático medio (RMSE)

Es una métrica que evalúa la precisión de los modelos de calibración al cuantificar la diferencia promedio entre los valores arrojados por el modelo y los valores reales, midiendo su capacidad predictiva. Un valor RMSE menor significa que el modelo es preciso al haber menor desviación entre los valores predichos y los verdaderos (Christin Brettschneider et al., 2022). Se calcula con la siguiente ecuación:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Donde:

y_i son los valores reales.

\hat{y}_i son los valores predichos por el modelo.

n es el número total de muestras.

2.2.6.4. Coeficiente de determinación (R^2)

Es una métrica que indica la proporción de variación en la variable dependiente explicada por los predictores incluidos en un modelo de regresión lineal (Zhang, 2017). Se emplea una función de varianza destinada a definir la variación total de la variable dependiente y la variación restante de la variable dependiente después de modelar los efectos predictivos de las variables independientes (Zhang, 2017).

Figura 4

Fórmula para calcular el coeficiente de determinación.

$$\begin{aligned} \text{Coeficiente de Determinación} &\rightarrow R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \\ \text{Suma de Cuadrados Total} &\rightarrow SST = \sum (y - \bar{y})^2 \\ \text{Regresión de Suma de Cuadrados} &\rightarrow SSR = \sum (y' - \bar{y})^2 \\ \text{Error de Suma de Cuadrados} &\rightarrow SSE = \sum (y - y')^2 \end{aligned}$$

Nota. Tomada de *Machine Learning Supervisado: Fundamentos de la Regresión Lineal*, por Roman, 2019, Medium, <https://medium.com/datos-y-ciencia/machine-learning-supervisado-fundamentos-de-la-regresi%C3%B3n-lineal-bbcb07fe7fd>.

Un R^2 cercano a 1 indica que es un buen modelo, no obstante, si es 1 puede significar que existe un sobreajuste. Por ende, el modelo no debería llegar a un coeficiente de determinación demasiado alto.

2.3. Marco Conceptual

2.3.1. Espectroscopía Infrarroja por Transformada de Fourier (FTIR)

Es una técnica analítica que determina los grupos funcionales de una sustancia, junto con los posibles enlaces moleculares entre compuestos químicos (Mohamed et al., 2017). Es aplicable para un gran abanico de compuestos, por lo que se empleará para obtener los espectros de las soluciones azucaradas y modelar sus concentraciones.

2.3.2. Glucosa

Es un monosacárido miembro del grupo de las aldohexosas y uno de los ingredientes principales de alimentos y bebidas. La glucosa se obtiene, en mayor medida,

a partir de lo que ingerimos, proveyendo energía a todos los sistemas complejos del organismo (Koirala et al., 2022).

2.3.3. Fructosa

Es un monosacárido de tipo cetohexosa ampliamente usado en la industria alimentaria. No obstante, no está presente en abundancia en la naturaleza, por lo que se genera a gran escala en la producción de sirope de maíz, un endulzante con alto contenido de fructosa (Lara-Serrano et al., 2021). Es un analito peculiar por su comportamiento espectral, atribuido a la presencia del grupo cetónico.

2.3.4. Sacarosa

Es un disacárido compuesto por glucosa y fructosa, y el edulcorante más popular. Dentro del organismo cumple una doble función: es un sustrato energético y una molécula de señalización (Lara-Cruz & Jaramillo-Botero, 2022).

2.3.5. Modelo matemático multivariado

Se usa para "predeterminar en base a una estandarización de diferentes variables, primeramente, definiendo la agrupación independiente de una variable que se puede predecir, respecto a una variable de respuesta" (Jiménez et al., 2022). Constituye el eje central del estudio ya que busca predecir los valores de los azúcares a partir de los espectros.

2.3.6. Regresión por mínimos cuadrados parciales

Es un método estadístico del análisis multivariado que busca establecer la relación entre los cientos de variables (números de onda) que proveen los datos espectroscópicos (Matienzo, 2021).

2.3.7. Preprocesamiento espectral

Son procedimientos que mejoran la calidad del modelo al eliminar bandas ruidosas o no calibradas y reducir la dimensionalidad de los datos (Orozco-Morales et al., 2025), mejorando las predicciones del modelo y resaltando las señales de interés de los azúcares.

2.3.8. Validación del modelo

Consiste en evaluar la capacidad de predicción del modelo, garantizando que sea confiable y reproducible en muestras reales de bebidas. Normalmente se realizan pruebas de normalidad, estabilidad e independencia (Jiménez et al., 2022).

CAPÍTULO III

MARCO METODOLÓGICO

3.1. Nivel de Investigación

Según la clasificación descrita por Arias (2012), la presente investigación corresponde al nivel de investigación explicativo, debido a que se plantea determinar la relación causa-efecto de las variables independientes (longitudes de onda) sobre la variable dependiente (concentración de los azúcares).

3.2. Diseño de Investigación

El diseño de investigación es experimental, ya que las muestras se analizarán a diversas concentraciones controladas para poder evidenciar la relación causa-efecto. También, es un diseño documental, al haber profundizado sobre las temáticas y conceptos del estudio a través de una revisión bibliográfica en bases de datos confiables.

3.3. Diseño de Experimento

Se desarrollará un diseño de mezclas con restricciones para generar la base de datos de los espectros y así entrenar los modelos. Particularmente, se crearán los estándares con distintas concentraciones de cada carbohidrato según los valores que, según las recomendaciones, deberían tener las bebidas azucaradas comerciales. Además, para acercarse a escenarios reales y extender el rango de calibración del modelo, se trabajarán con concentraciones variables a las recomendadas por entidades regulatorias.

Las concentraciones de azúcares presentes en las bebidas azucaradas se expresarán en % p/v para su análisis. Por otro lado, para generar las mezclas bajo las restricciones establecidas, se utilizará el software RStudio junto con los paquetes especializados *rms* y *dplyr*.

A continuación, se muestra la tabla con el diseño de mezclas y su respectivo código. Se trabajarán con 32 muestras para la calibración del modelo y 3 muestras comerciales para la validación del modelo. En total se analizarán 35 muestras y cada muestra tendrá 3 réplicas.

Tabla 3

Mezclas del diseño experimental destinadas a la calibración del modelo

Código	Glucosa % p/v	Fructosa % p/v	Sacarosa % p/v	Agua (%)
M1	6.00	0.00	0.00	94.00
M2	0.00	6.00	0.00	94.00
M3	0.00	0.00	6.00	94.00
M4	3.00	3.00	0.00	94.00
M5	3.00	0.00	3.00	94.00
M6	0.00	3.00	3.00	94.00
M7	2.00	2.00	2.00	94.00
M8	4.50	1.50	0.00	94.00
M9	4.50	0.00	1.50	94.00
M10	0.00	4.50	1.50	94.00
M11	1.50	4.50	0.00	94.00
M12	1.50	0.00	4.50	94.00
M13	0.00	1.50	4.50	94.00
M14	2.25	2.25	1.50	94.00
M15	2.25	1.50	2.25	94.00
M16	1.50	2.25	2.25	94.00
M17	3.75	1.00	1.25	94.00
M18	3.75	1.25	1.00	94.00
M19	1.25	3.75	1.00	94.00
M20	1.00	3.75	1.25	94.00
M21	1.25	1.00	3.75	94.00

M22	1.00	1.25	3.75	94.00
M23	2.50	2.50	1.00	94.00
M24	2.50	1.00	2.50	94.00
M25	1.00	2.50	2.50	94.00
M26	3.00	2.00	1.00	94.00
M27	1.00	3.00	2.00	94.00
M28	0.5	0.5	0.5	98.50
M29	1.00	1.00	1.00	97.00
M30	1.50	1.50	1.50	95.50
M31	2.00	2.00	2.00	94.00
M32	2.50	2.50	2.50	92.50

Nota. Propia autoría.

3.4. Técnicas e Instrumentos de Recolección de Datos

Las técnicas de recolección de datos empleadas en la presente investigación son el análisis documental y el análisis de contenido de diversas fuentes fidedignas como libros y artículos científicos publicados en bases de datos académicas como Scopus, Elsevier, PubMed, SciELO y Google Académico.

Adicionalmente, se llevó a cabo una observación estructurada de los espectros provenientes del equipo de espectroscopía FTIR, los cuales se almacenarán en archivos con formato .CSV para su posterior análisis en el software RStudio.

3.5. Técnicas de Procesamiento y Análisis de Datos

En primer lugar, el preprocesamiento de datos se ejecutó en el software RStudio a través de técnicas de correlación de línea base (ALS), suavización (SG), normalización de datos, primera derivada y segunda derivada. Luego, se aplicaron métodos de regresión multidimensional y reducción de dimensionalidad, entre los que se encuentran el análisis de componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS).

3.6. Protocolo a implementar

3.6.1. Diseño y preparación de mezclas

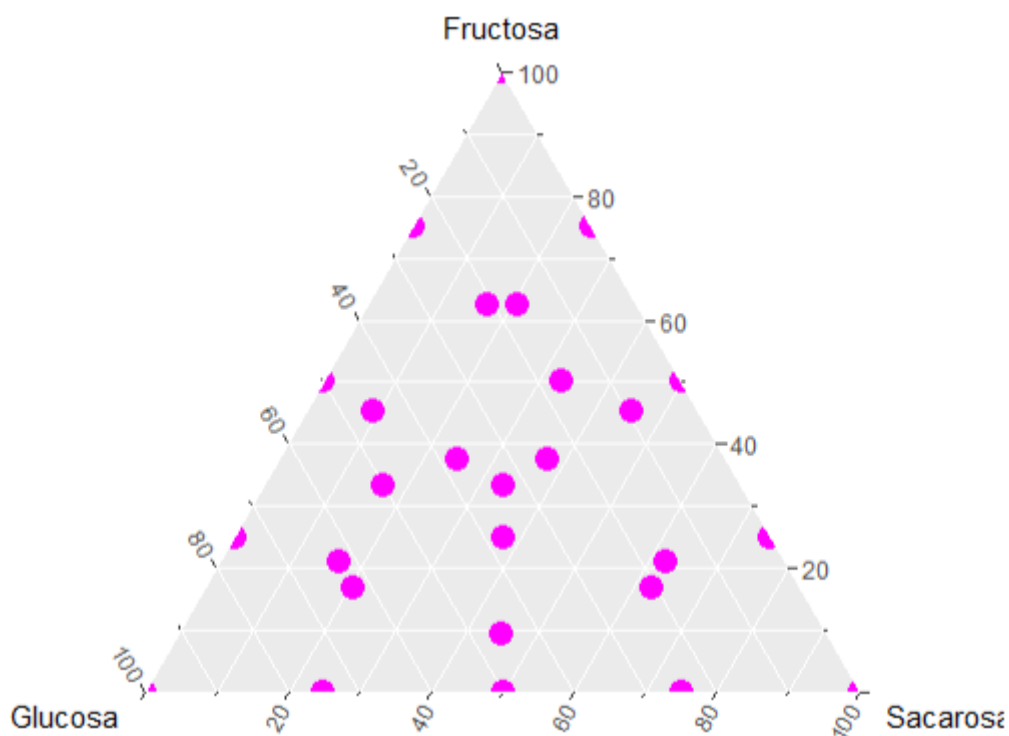
El diseño experimental establece restricciones según las recomendaciones de la Organización Mundial de la Salud, una concentración de azúcares que no supere el 6%. Otras muestras comprenden concentraciones mayores y menores al 6%.

Para la preparación de las mezclas, se emplearon sustancias de grado analítico y se siguieron normas de bioseguridad para evitar la contaminación. En primer lugar, se realizaron diluciones de los tres azúcares, para luego preparar las mezclas; obteniéndose un total de 32 muestras destinadas al desarrollo del modelo.

Para una mejor interpretación de los datos, se generó una gráfica en RStudio, de dispersión ternaria con el paquete *ggtern*.

Figura 5

Gráfico de dispersión ternaria de las mezclas



Nota. Propia autoría

3.6.2. Corrida de las muestras en el equipo FTIR-ATR

3.6.2.1. Obtención del espectro de fondo

Previo a la corrida de las muestras, se realizó el análisis del *background* para asegurar la precisión del equipo. Este análisis se repitió cada veinte minutos durante todo el proceso.

3.6.2.2. Preparación y medición de muestras en el equipo.

Las mezclas se almacenaron en tubos Eppendorf de 2 mL, y a cada muestra se le asignó un código para identificar cada réplica y el número de muestra. Luego, se limpió cuidadosamente la punta de diamante con etanol al 70%. Por último, se homogeneizaron las muestras y se ejecutó el análisis colocando 55 μL en la zona de medición.

3.6.2.3. Almacenamiento de espectros

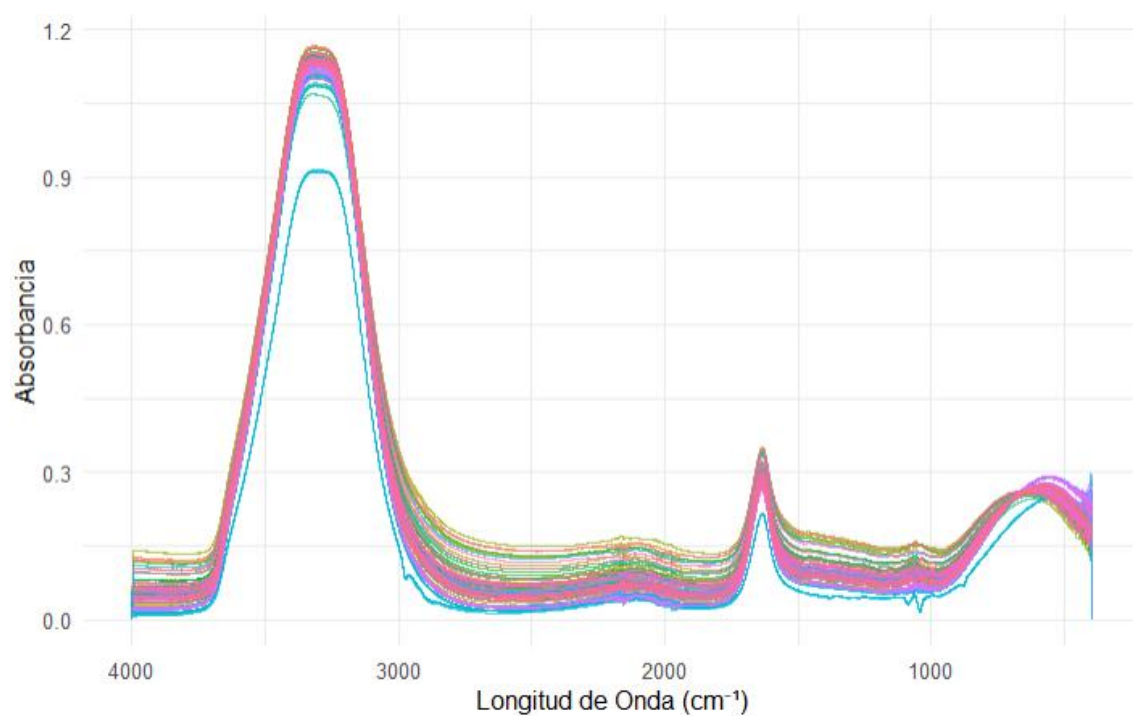
Los datos de las mezclas se guardaron en forma cruda en formato .CSV. Es decir, sin someterlos a modificaciones en el software OMNIC, para luego ser importados en RStudio usando las librerías *ir* y *dplyr*. Cabe recalcar que se generaron dos bases de datos, la primera con el espectro completo y la segunda con el espectro recortado.

Previo al desarrollo del modelo, se debe establecer si se trabajará con el espectro completo, como se muestra en la Figura 5, o con el espectro recortado, representado en la Figura 6; aquello mediante la revisión de la literatura y la verificación de regiones que tengan alta variabilidad sistemática. Para procesar los datos destinados al recorte espectral, se estableció que se abarcará la región de la huella digital de los carbohidratos, comprendida entre 850 y 1200 cm^{-1} , ya que los analitos (glucosa, fructosa y sacarosa) poseen espectros similares. Entonces, existe la posibilidad de que si se trabaja con la zona de la huella digital se podrían identificar diferencias estructurales sutiles que maximicen

la información relevante para el modelo y descarten señales innecesarias y ruido espectral. Así, se evita el sobreajuste asociado al solapamiento espectral y mejora la capacidad predictiva del modelo PLS. Los datos correspondientes al espectro completo se mantuvieron sin ninguna modificación.

Figura 6

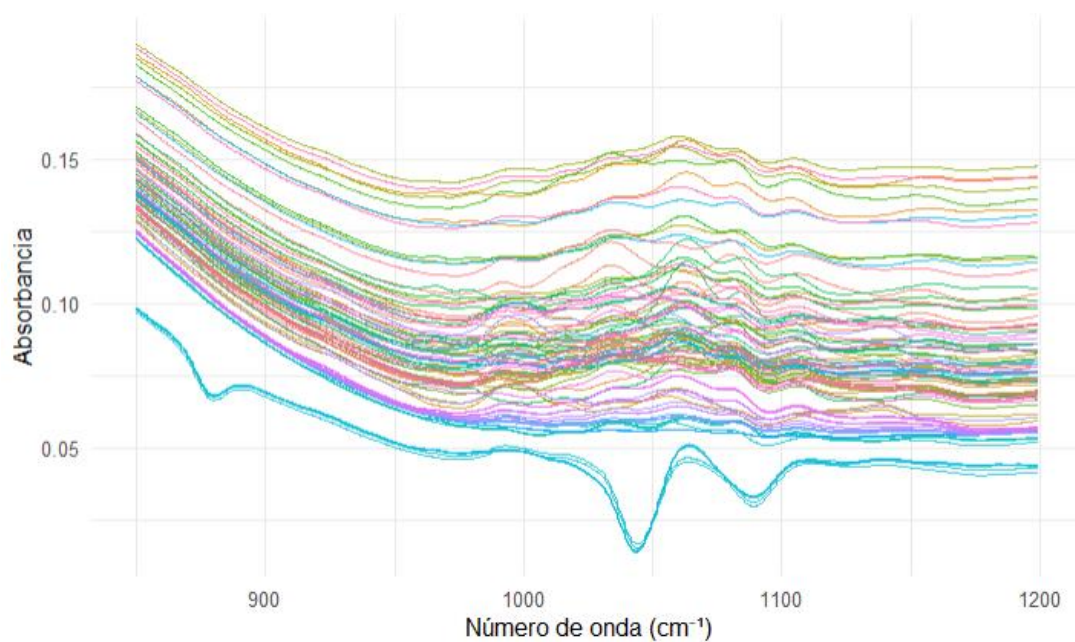
Espectros completos de las mezclas.



Nota. Se muestra el espectro completo de las muestras, desde la región de 4000 cm⁻¹ hasta 500 cm⁻¹. Propia autoría.

Figura 7

Espectros después del recorte.



Nota. Propia autoría

3.6.2.4. Preprocesamiento

Los datos crudos del espectro completo y del espectro recortado pasaron por una fase de preprocesamiento, la cual consiste en la primera derivada y la segunda derivada, dentro de las cuales se llevó a cabo la normalización, corrección de la línea base y suavizado.

3.6.2.4.1. Primera derivada

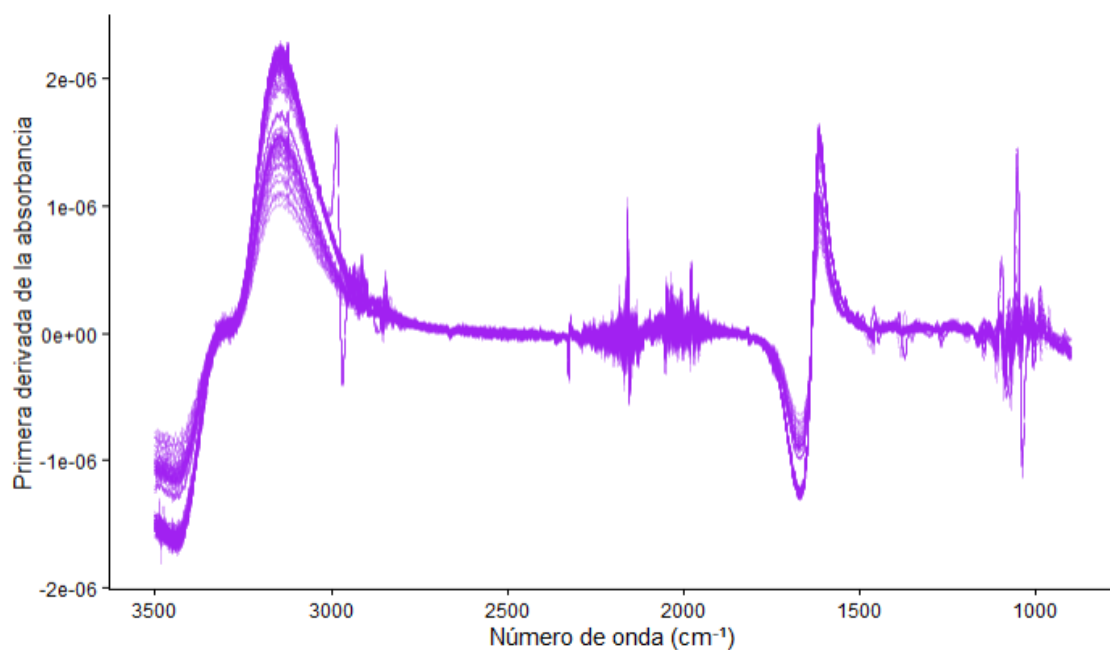
La primera derivada elimina la línea base inclinada, aunque para cuantificación es necesario también aplicar un espectro de segunda derivada (Frost, 1999). Además, resalta diferencias químicas sutiles entre analitos que comparten similitudes estructurales (Rosales-Martínez et al., 2025).

3.6.2.4.2. Normalización

Busca minimizar variaciones de intensidad no asociadas a la concentración de los azúcares. Para ello, se divide cada valor de intensidad por la suma de los valores de intensidad. En R, se empleó la función "ir_normalise".

Figura 8

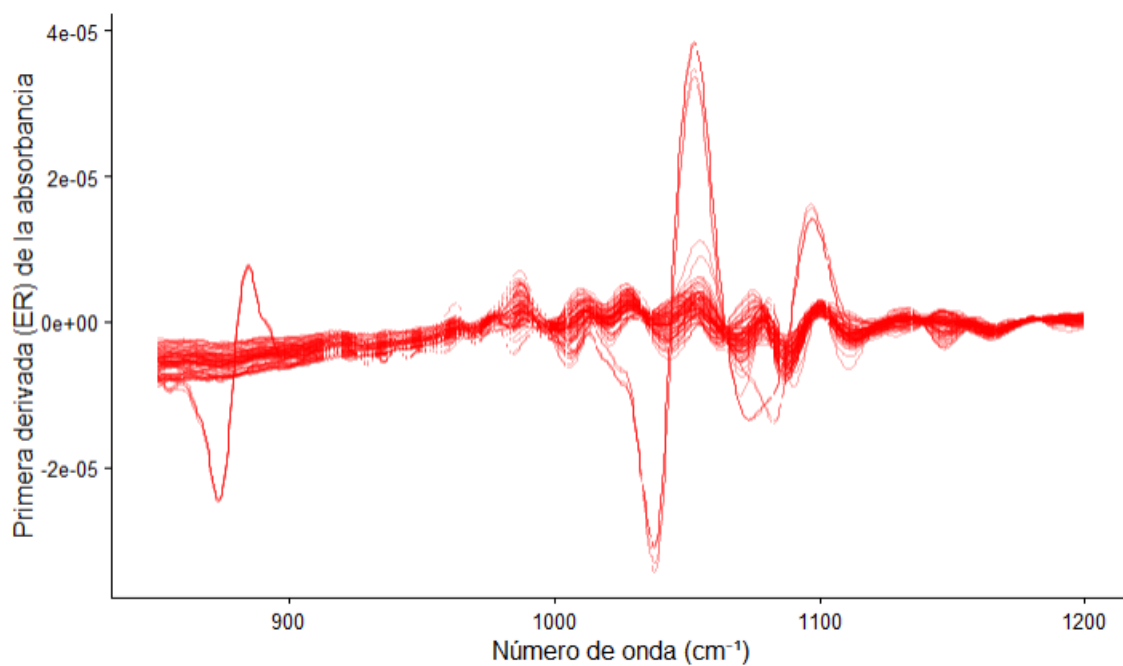
Espectros completos luego de aplicarse la primera derivada



Nota. Propia autoría.

Figura 9

Espectros recortados luego de aplicarse la primera derivada



Nota. Propia autoría.

3.6.2.4.3. Segunda derivada

La segunda derivada se encargará de eliminar cualquier curvatura de la línea base, siempre y cuando se ajuste a una ecuación cuadrática con respecto a la longitud de onda (Frost, 1999).

3.6.2.4.4. Corrección de la línea base

En cuanto a la corrección de la línea base, se empleó el método ALS, el cual consiste en una penalización asimétrica que elimina interferencias y minimiza la presencia de la línea base sin alterar los picos relevantes (Peng et al., 2010). Para este paso, se utilizó la función *baseline* en la segunda derivada.

3.6.2.4.5. Suavizado

Se utilizó el método de Savitzky-Golay, a fin de minimizar el ruido espectral y calcular las derivadas al ajustar una curva polinómica dentro de una ventana. Así, mejora la resolución espectral y, por ende, la precisión del modelo de regresión multivariante (Delwiche & Reeves, 2010).

En el suavizado se consideran las siguientes variables descritas por Savitzky & Golay (1964):

- p es el orden del filtro, descrito como el grado del polinomio empleado en el ajuste.
- n es la cantidad de números de onda necesarios para construir el polinomio.
- m es la derivada que se pretende calcular.

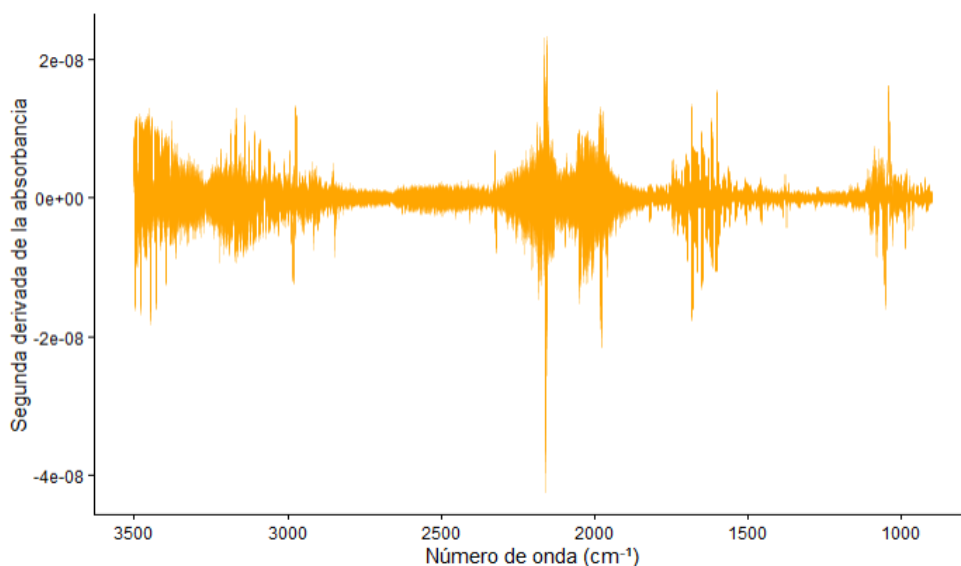
Para la primera derivada se utilizaron las siguientes combinaciones ($p = 2, n = 11, m = 0$; $p = 2, n = 11, m = 1$). Primero, se aplicó un suavizado sin derivada ($m = 0$) para

disminuir el ruido instrumental y mejorar la estabilidad. Luego, el polinomio de segundo orden explica la forma local de las bandas sin llegar al sobreajuste, y una ventana de 11 puntos reduce el ruido, pero preserva la información.

Mientras que, para la segunda derivada, se empleó ($p = 2$, $n = 13$, $m = 2$), una ventana mayor para controlar la amplificación del ruido, al ser una segunda derivada; resaltar bandas solapadas; y eliminar residuos de la línea base.

Figura 10

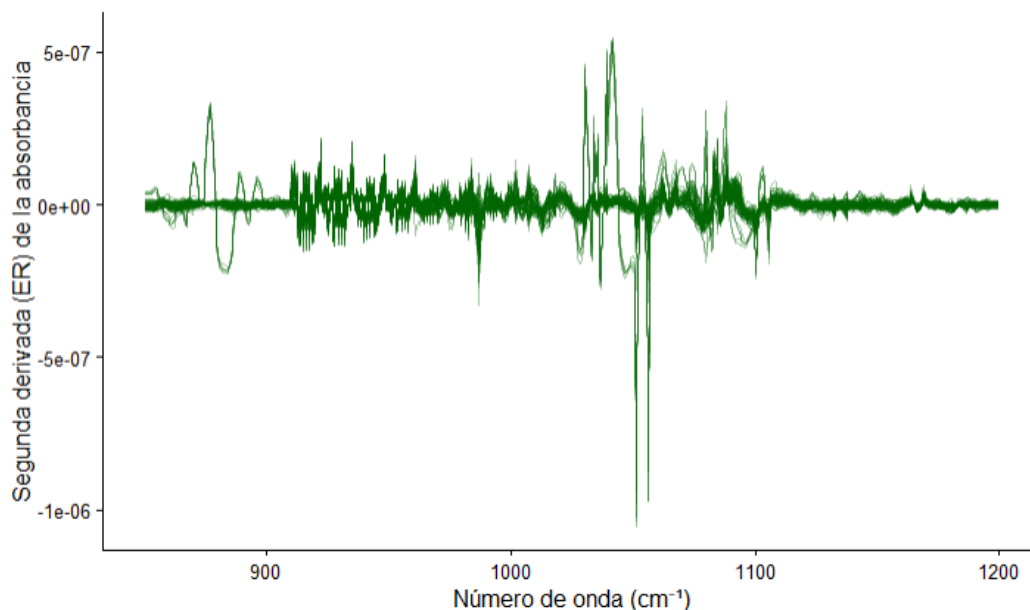
Espectros completos luego de la segunda derivada



Nota. Propia autoría.

Figura 11

Espectros recortados luego de la segunda derivada



Nota. Propia autoría.

3.6.3. Métodos estadísticos

3.6.3.1. Análisis de componentes principales (PCA)

Este análisis se ejecutó en la primera y segunda derivada de los dos grupos de datos, donde en la primera derivada contribuyó a la evaluación de la estructura de los datos y los patrones asociados a las distintas concentraciones. Por otro lado, se aplicó en la segunda derivada para verificar que, después del preprocesado, la información química esencial se haya conservado.

3.6.3.2. Entrenamiento y evaluación del modelo

El conjunto de datos se dividió en dos partes de forma aleatoria, un 70% de las variables se destinó al entrenamiento del modelo, y el 30% restante se estableció para la prueba de este. Este rango se estableció considerando la limitada cantidad de muestras, por lo que se prioriza el entrenamiento. Además, se garantizó que las réplicas de cada muestra permanezcan juntas, para evitar la sobreestimación del desempeño del modelo ya que las réplicas no son muestras independientes.

3.6.3.3. Regresión por mínimos cuadrados parciales (PLS)

Este método modela la relación de los espectros preprocesados con las concentraciones de los analitos, basándose en la colinealidad significativa de las variables y el número de predictores relacionados con el número de muestras. Para el desarrollo de la regresión, se estableció la variable “X” como las longitudes de onda; mientras que “Y” son las concentraciones de glucosa, fructosa y sacarosa.

3.6.3.4. Error cuadrático medio (RMSE)

Una vez desarrollado el modelo, se evaluó su rendimiento calculando el error cuadrático medio. Luego, se ordenaron los resultados en un formato de lista donde se muestra el analito con su respectivo valor de error cuadrático medio.

3.6.3.5. Coeficiente de determinación (R^2)

El coeficiente de determinación explica qué proporción de la variabilidad de Y se puede explicar con el modelo de regresión desarrollado.

3.6.3.6. Validación del modelo

Para evaluar la capacidad predictiva, se estimará la precisión con la cual el modelo pueda identificar la concentración de azúcares de 3 muestras comerciales.

En primer lugar, se obtuvieron los espectros de las muestras con sus respectivas réplicas y códigos para ser identificadas. Luego, se importan los datos a R y los espectros fueron analizados por el modelo de regresión, evaluándose los resultados de las métricas del RMSE y R^2 . A una de las muestras comerciales se le realizó un análisis fisicoquímico por cromatografía líquida de alto rendimiento (HPLC) para obtener el perfil de azúcares, y así comparar qué tan eficaz es el modelo conociendo la verdadera concentración de los azúcares de la bebida.

CAPÍTULO IV

ANÁLISIS DE RESULTADOS Y DISCUSIÓN

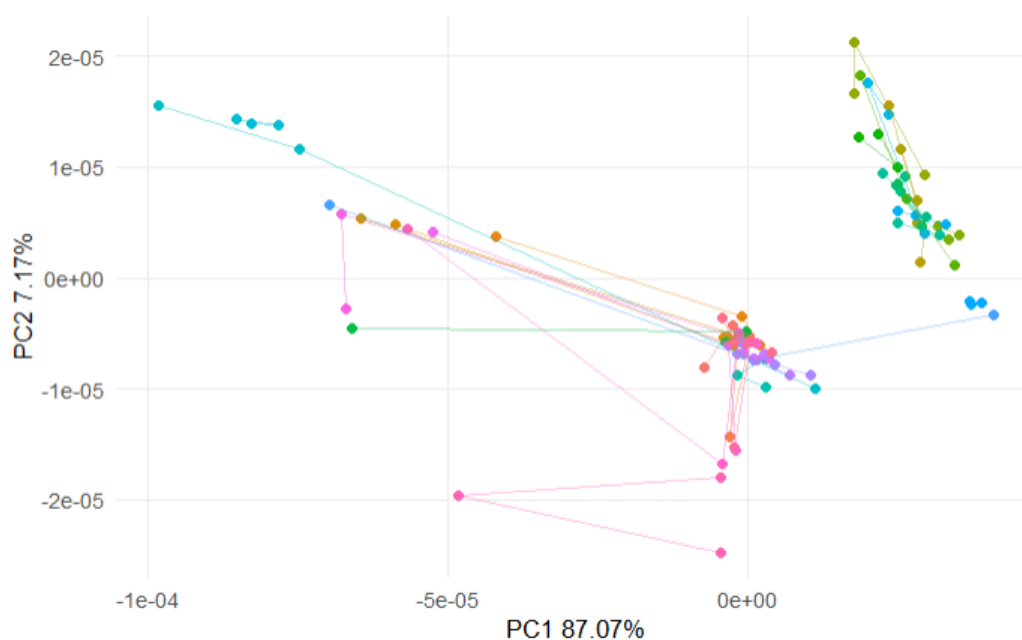
4.1. Resultados experimentales obtenidos

4.1.1. Análisis de componentes principales

En el análisis de componentes principales se trabajó con la primera y segunda derivada, considerando cada mezcla como una unidad experimental, por lo que se agruparon sus réplicas correspondientes.

Figura 12

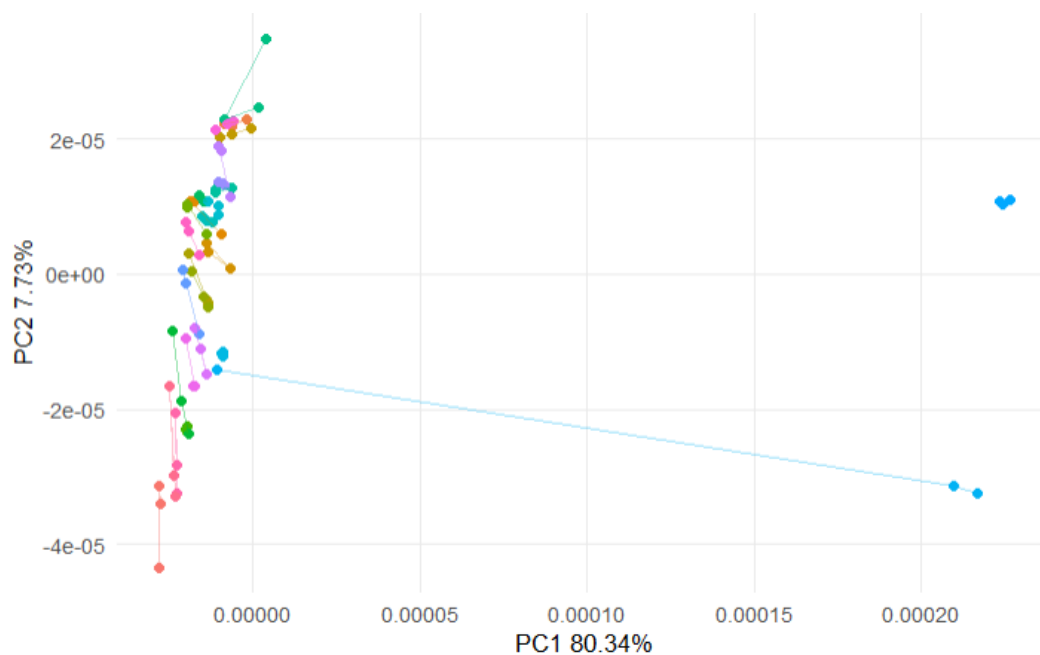
Gráfica del análisis de componentes principales de la primera derivada de los espectros completos.



Nota. Propia autoría.

Figura 12

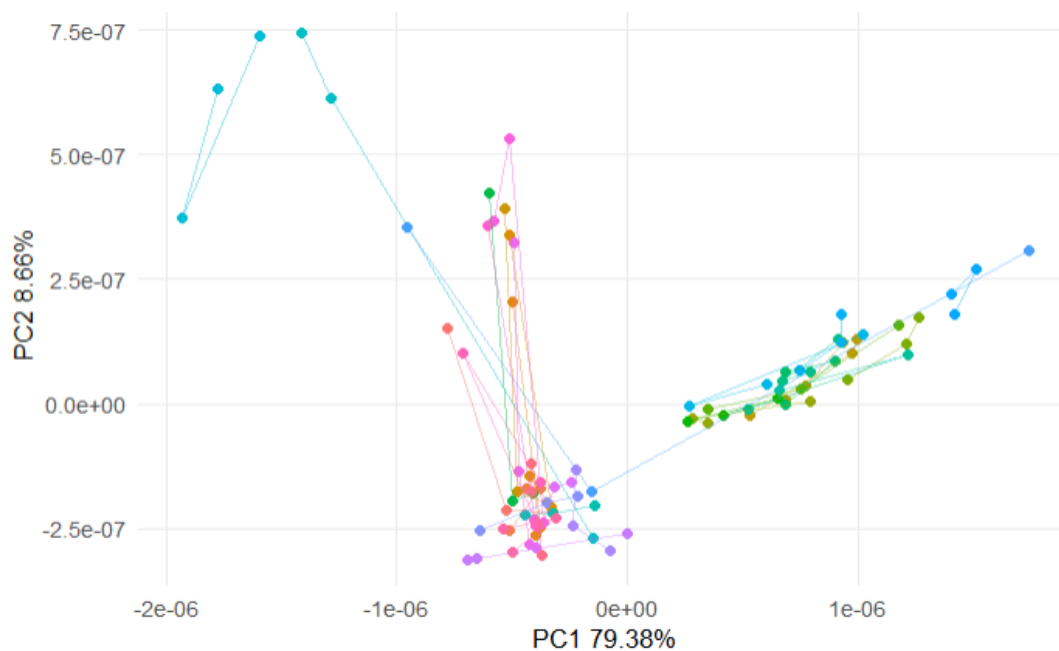
Gráfica del análisis de componentes principales de la primera derivada de los espectros recortados.



Nota. Propia autoría.

Figura 13

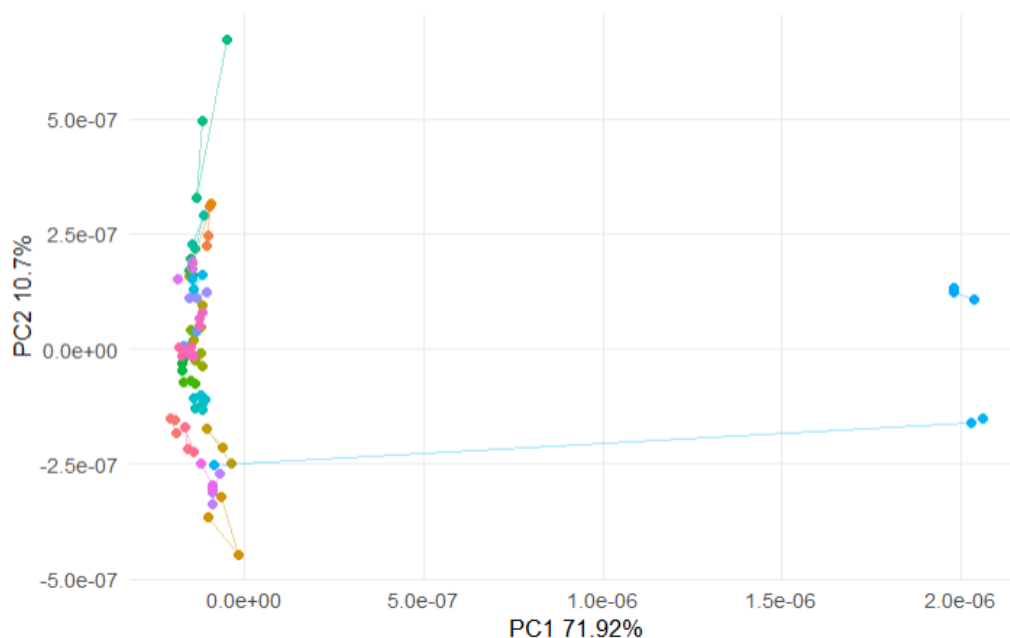
Gráfica del análisis de componentes principales de la segunda derivada de los espectros completos.



Nota. Propia autoría.

Figura 14

Gráfica del análisis de componentes principales de la segunda derivada de los espectros recortados.



Nota. Propia autoría.

El análisis de componentes principales (PCA) indica que, en la primera derivada de los espectros completos, el primer componente explica el 87.07 % de la varianza total y el segundo componente explica el 7.17 % de la varianza espectral. Es decir, en el primer componente la mayoría de la información espectral se relaciona a un gradiente químico asociado con la composición de azúcares. Por otro lado, el segundo componente encontró diferencias sutiles entre cada mezcla. Para la segunda derivada, el primer componente explica el 79.38 % de la varianza total y el segundo componente explica el 8.66 % de la varianza espectral, evidenciando que hay una redistribución de la varianza dirigida a componentes superiores por las diferencias espectrales en la forma y posición de bandas específicas de los azúcares.

En los espectros recortados se observó algo similar, el primer componente de la primera derivada explica el 80.34 % de la varianza total, y el segundo componente explica el 7.73 % de la varianza espectral. Mientras que, para la segunda derivada, el primer componente explica el 71.92 % de la varianza total y el segundo componente explica el 10.7 % de la varianza espectral. La segunda derivada demostró menor varianza en el primer componente y una contribución superior en los siguientes componentes; es decir, denotando características espectrales específicas.

Adicionalmente, en las gráficas de los espectros recortados se puede visualizar proximidad entre las réplicas, lo cual indica que el preprocesamiento fue consistente y que existe repetibilidad del análisis.

4.1.2. Selección de rangos óptimos

En primer lugar, desde una perspectiva química, existen regiones en las cuales las vibraciones de ciertos grupos funcionales predominan. En el caso de los carbohidratos, tales regiones son vibraciones de estiramiento y deformación de enlaces C–O y C–C, encontradas en la región de la huella digital. Por otro lado, la región diagnóstica del espectro muestra solapamiento, absorción por las bandas del disolvente y ruido, por lo que son poco informativas. La selección del rango espectral debe enfocarse en regiones que provean absorciones características de los analitos de interés, por lo que en quimiometría es común no considerar la región diagnóstica en los análisis (Coldea et al., 2013).

Tras el preprocesamiento de todo el espectro, después de la primera derivada, en la Figura 7 se muestra que hay variables espectrales con valores faltantes o con alta variabilidad en todo el rango espectral. Es decir, a pesar de que la primera derivada mejora la estructura, en el espectro completo aún se enfatizan regiones con vibraciones del grupo

O-H del disolvente y ruido; aquello dificulta identificar zonas con información química relevante. Luego de la segunda derivada, se resaltan picos de interés, pero se observa que las señales se vuelven inestables en regiones de baja absorbancia y se amplifica el ruido.

En cuanto al preprocesamiento del espectro recortado, en la primera derivada hay menor dispersión entre las muestras, se diferencian patrones y existe menor ruido de fondo. De manera similar, en la segunda derivada, las bandas vibracionales correspondientes a los carbohidratos son perceptibles y existe menor ruido.

Por otro lado, en el análisis de componentes principales (PCA) de las primeras y segundas derivadas, se evidencia que una porción significativa de la varianza espectral está concentrada en pocos componentes, aludiendo a que no todas las regiones del espectro aportan variabilidad significativa. Además, las gráficas de score indican que las muestras tienen una tendencia a agruparse cercanamente en los espectros recortados, diferenciando de mejor forma la variación química de los azúcares.

En base a la interpretación expuesta, se seleccionó el rango espectral comprendido entre 850 y 1200 cm^{-1} para la construcción del modelo.

4.1.3. Modelo de regresión por mínimos cuadrados parciales (PLS)

En R, el modelo fue construido bajo los siguientes parámetros:

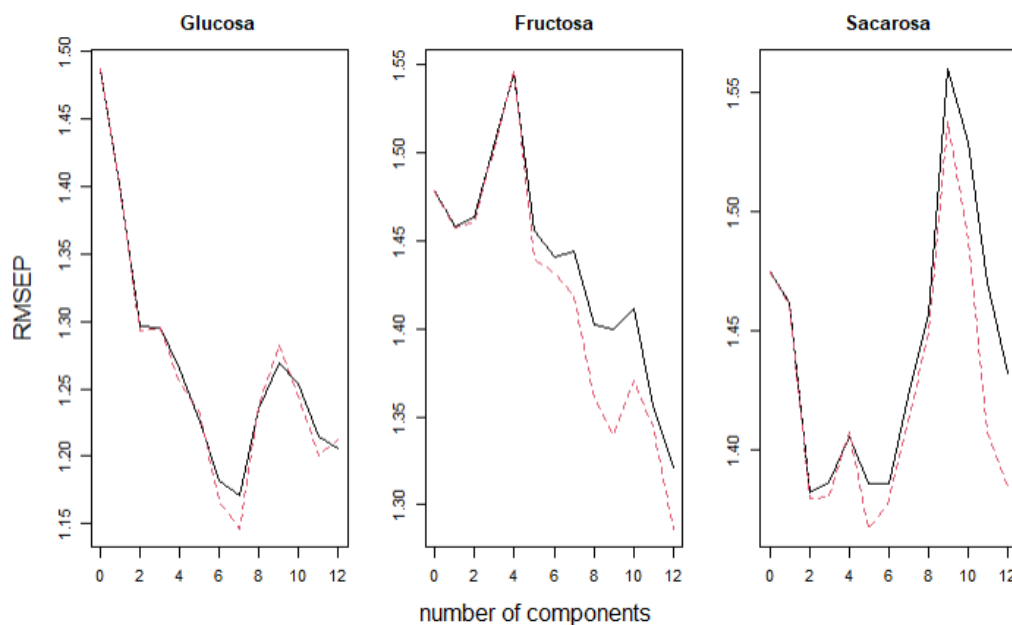
- Tipo de modelo: Regresión
- Número de componentes: 12
- Dimensión de los datos

X (predictores): 96 – 814 (96 muestras y 814 variables espectrales).

Y (variable de respuesta): 96 – 3 (96 muestras y 3 variables de respuesta).

Figura 15

Evolución del error cuadrático medio de predicción (RMSEP) en función del número de componentes del modelo.



Nota. Propia autoría.

Las gráficas demuestran que con doce componentes se minimiza el error de predicción para la glucosa y fructosa, mientras que para la sacarosa el número óptimo de componentes es seis. Por ende, para el modelo se trabajó con doce componentes.

El modelo se basa en una gran cantidad de variables espectrales, y la ecuación implica una combinación lineal de todos los números de onda con sus coeficientes, los cuales pueden ser extraídos en el software. No obstante, debido a la gran cantidad de coeficientes, se mostrarán solo los más influyentes.

Tabla 4

Coefficientes más influyentes en el modelo con sus respectivos números de onda.

Número de onda	Coefficiente
850.0076	176.748
850.4897	176.748

850.9718	176.748
851.454	176.748
851.9361	176.748
852.4182	176.748
1005.738	150284.0
1004.774	149926.5
998.9879	-147959.5
1005.256	145755.6
999.47	-144493.9
1004.773	144248.4
1080.469	131940.5
1004.291	129713.6
1006.22	128795.8
1055.398	-128794.7
1055.88	-126622.0
999.47	-126231.8
1055.398	-125534.3
1055.88	-125307.1

Nota. Propia autoría.

Las concentraciones aproximadas de cada analito que el modelo predijo de las muestras de prueba se muestran en la siguiente tabla:

Tabla 5

Predicción de las muestras del entrenamiento.

Muestra	Réplica	Glucosa	Fructosa	Sacarosa
1	1	5.66335197	-0.1673195	0.84247240
	2	5.17305821	1.0424055	0.17941315
	3	5.14889545	0.1015048	0.70492025
2	1	0.27233386	5.9910247	0.04943535
	2	-0.27897177	5.8934405	0.43991644
	3	0.56202175	4.7294590	0.70379587
4	1	1.88528803	2.7207599	1.04877892
	2	3.67937909	1.7633834	0.24708409
	3	3.13510502	2.4972787	0.38714670
6	1	1.17875664	2.7926978	2.39795159
	2	1.30562132	1.8017534	3.22652722
	3	1.52654244	2.4430344	2.39978918
12	1	1.53465877	-0.6599989	5.35792324
	2	1.90327787	0.4828734	3.91446336
	3	0.91040700	-0.4051847	4.51952917
	1	0.17439038	2.4256448	3.29458856

13	2	-0.09350889	2.9389466	2.63647563
	3	-0.17471518	2.6859209	3.14720922
	1	0.46211887	2.1810126	2.58100361
16	2	1.36143205	2.8632204	1.28890458
	3	1.05755145	1.5164172	2.85863791
	1	1.21871114	0.3637143	4.78907927
21	2	1.19803867	-0.0267652	4.68514280
	3	1.90639085	1.1411740	3.31205370
	1	1.32354614	0.8974207	3.08855996
24	2	2.44892669	1.9970119	2.42910695
	3	3.29474163	1.2622772	2.12414528
	1	1.82735113	1.8302475	2.63136746
25	2	0.91971537	2.8940341	3.14960546
	3	1.75951572	1.2826935	3.29558451

Nota. Los valores están expresados en %m/v. Propia autoría.

4.1.4. Error cuadrático medio (RMSE) y coeficiente de determinación (R^2)

En la siguiente tabla se detallan los resultados del error cuadrático medio y coeficiente de determinación del modelo.

Tabla 6

Error cuadrático medio y coeficiente de determinación calculados a partir del modelo.

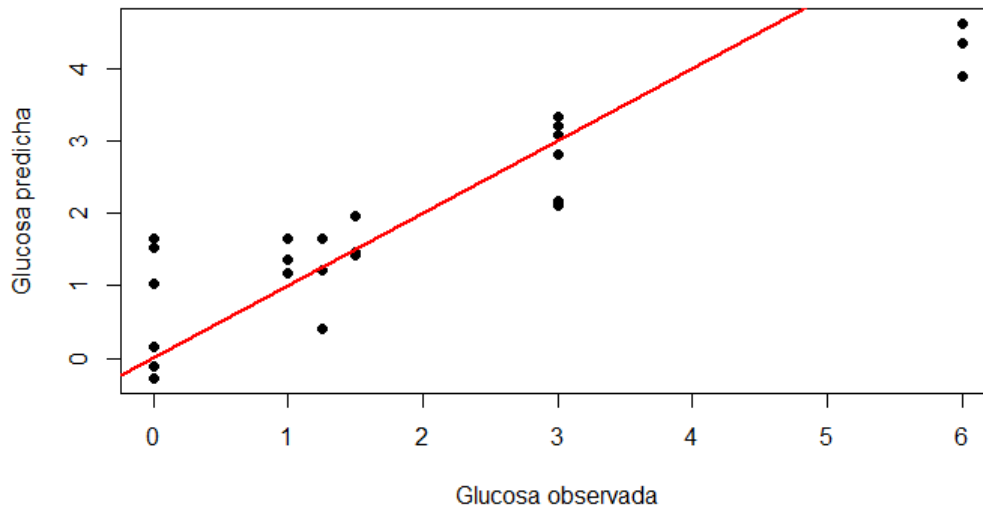
Analito	RMSE	R^2
Glucosa	0.7539088	0.8195269
Fructosa	0.8077523	0.7820279
Sacarosa	0.7411921	0.8071457

Nota. Propia autoría.

Para representar los resultados, se establecieron gráficas de observado versus predicho para comparar los valores predichos por el modelo con los valores observados para cada uno de los azúcares. Si los puntos se alinean con la diagonal, significa que el modelo es preciso; pero, si se dispersan, indica que existen errores en las predicciones.

Figura 16

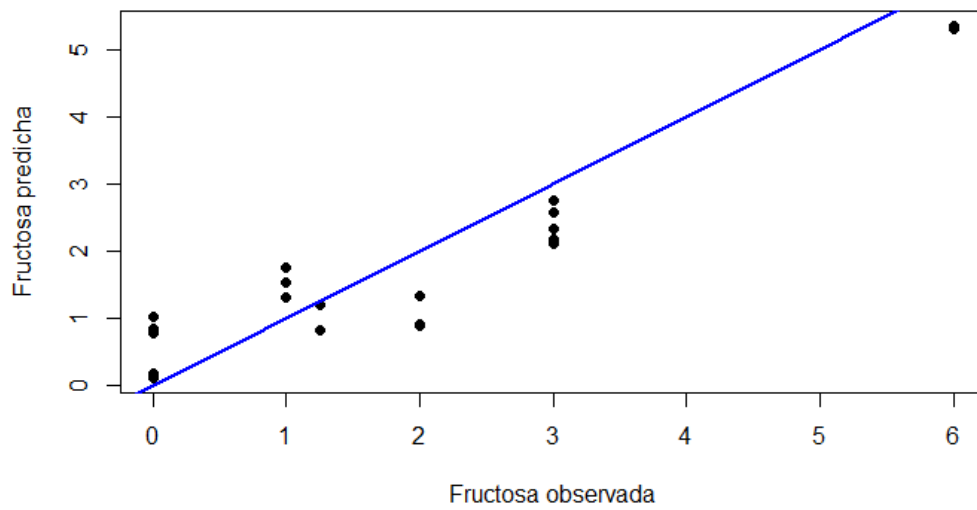
Gráfica de observado versus predicho, correspondiente a la glucosa.



Nota. Propia autoría.

Figura 17

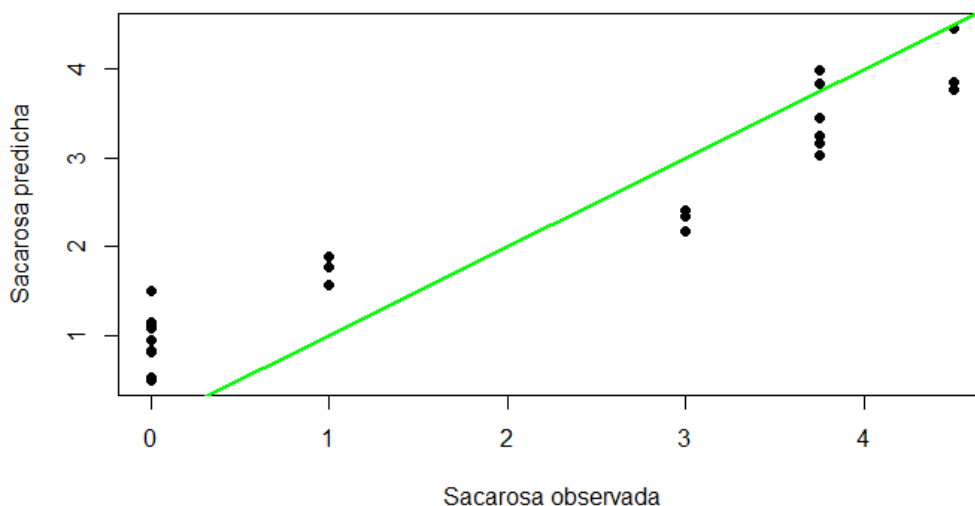
Gráfica de observado versus predicho, correspondiente a la fructosa.



Nota. Propia autoría.

Figura 18

Gráfica de observado versus predicho, correspondiente a la sacarosa.



Nota. Propia autoría.

4.1.5. Validación del modelo con muestras de bebidas comerciales

Se evaluó la capacidad predictiva del modelo con muestras de bebidas comerciales, obteniéndose concentraciones aproximadas de cada analito que se muestran en la siguiente tabla:

Tabla 7

Predicción de las muestras comerciales y el supuesto total de azúcares.

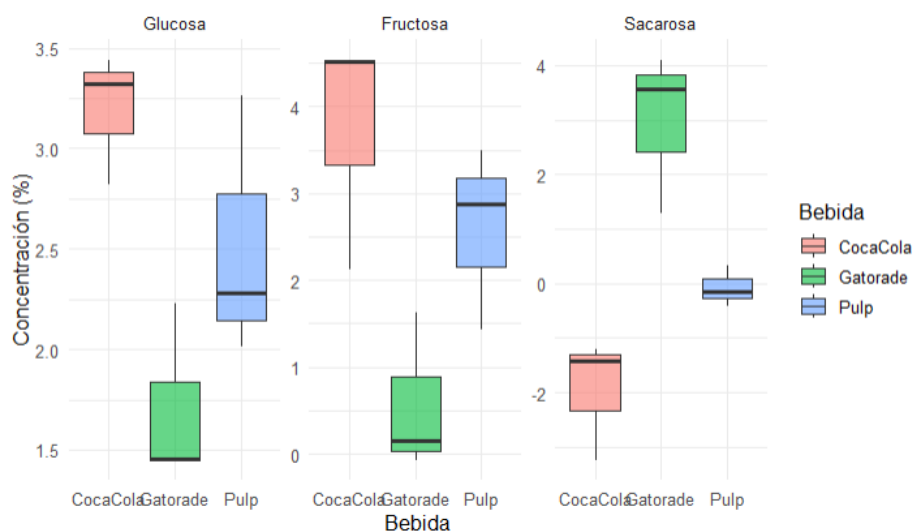
Bebida	Réplica	Glucosa	Fructosa	Sacarosa	Total
B1	1	2.823320	4.51391495	-1.4280448	5.909191
	2	3.320322	4.53958892	-3.2424207	4.617490
	3	3.441313	2.12895551	-1.1921232	4.378145
B2	1	2.229354	1.62476558	1.2784206	5.132541
	2	1.447642	-0.07318047	4.1061054	5.480567
	3	1.452832	0.15116356	3.5564718	5.160467
B3	1	3.260734	1.43397408	-0.1407421	4.553966
	2	2.282997	3.48967728	-0.4053495	5.367325
	3	2.010773	2.87159974	0.3196281	5.202001

Nota. Los valores están expresados en %v/v. Propia autoría.

Para visualizar cómo es la distribución de los azúcares en las bebidas comerciales, se representó con las siguientes gráficas:

Figura 19

Distribución de azúcares por bebida.



Nota. Propia autoría.

Finalmente, el análisis cuantitativo mediante HPLC arrojó los siguientes resultados:

Tabla 8

Perfil de azúcares de la bebida comercial.

Parámetro	Resultado	Método de análisis de referencia	Incertidumbre
Azúcares totales	9.80	AOAC 977.20 Modificado	-
Lactosa	0.00	AOAC 977.20 Modificado	-
Sacarosa	8.78	AOAC 977.20 Modificado	+/-1,12
Fructosa	1.02	AOAC 977.20 Modificado	+/-0,14
Glucosa	0.00	AOAC 977.20 Modificado	-

Nota. Los valores están expresados en %v/v.

4.2. Discusión

En primer lugar, para el preprocesamiento de los datos se aplicaron distintos métodos dentro de la primera y segunda derivada, como la corrección de la línea base mediante el ALS y *rubberband*, normalización, suavizado con el filtro Savitzky-Golay. En primera instancia, para la corrección de la línea base se probaron los dos métodos señalados anteriormente: el *rubberband* se estira sobre la curva de la línea base, y los puntos de esta que no se encuentren dentro del *rubberband* se eliminan, para luego restar los datos brutos con la línea base para obtener el espectro aplanado (Shen et al., 2018); por otro lado, el método de mínimos cuadrados penalizados (ALS) mejora la detección de las señales, ya que ajusta los datos pero asegura que se mantengan señales relevantes (Zhang et al., 2010).

Al analizar ambas metodologías, se escogió los mínimos cuadrados penalizados (ALS) ya que separa la línea base de señales con bandas anchas y solapadas, preservando la información. A comparación del método *rubberband*, el cual puede producir cierto grado de variabilidad por la selección de puntos de anclaje.

Respecto al suavizado, se ocupó el filtro de Savitzky-Golay tanto para la primera como para la segunda derivada, obteniendo buenos resultados. Para los modelos se empleó la segunda derivada, ya que la misma se calculó a partir de los espectros previamente procesados por la primera derivada, la cual elimina un desplazamiento de la línea base, pero es muy difícil de interpretar (Davies, 2007). Así, las derivadas de orden superior no son independientes, sino que también precisan de un tratamiento.

Ahora, el rango espectral óptimo resultó ser la zona entre 850 y 1200 cm^{-1} ; conocida por contener vibraciones de estiramiento y deformación de enlaces C–O, C–C y C–O–H (Leopold et al., 2011) de carbohidratos. Esto se debe a que, al incluir el recorte

espectral, el modelo solo aprende la información más relevante sobre los analitos, representando mejor la relación entre los datos espectrales y la variable de respuesta (Huang, 2022); reducir el ruido y variables irrelevantes (de Oliveira et al., 2026).

Considerando que los tres azúcares analizados poseen un espectro muy similar, la restricción minimiza la dimensionalidad de los datos, lo cual concuerda con las conclusiones de varios autores de que la restricción de variables espectrales mejora la capacidad predictiva de los modelos multivariados (Assis et al., 2019; Dhaulaniya et al., 2020; Sahlan et al., 2019) . También, Szaniawska et al. (2025) identificaron que, en la discriminación de muestras complejas por FTIR-ATR, las regiones espectrales más discriminatorias fueron la huella de carbohidratos, que mejoró la separación en los scores de PCA y permitió la diferenciación de patrones químicos. De igual manera, Huang (2022) respalda la idea de que las zonas de huellas espectrales infrarrojas pueden detectar diferencias químicas específicas, especialmente en las de azúcares.

Posteriormente, se estableció el modelo de regresión por mínimos cuadrados parciales (PLS) con el número de componentes que demostró proporcionar un menor error cuadrático medio, en este caso doce componentes. Las métricas del coeficiente de determinación (R^2) y error cuadrático medio (RMSE) resultantes en la fase de prueba, aluden a que el modelo tiene una capacidad de ajuste y predicción aceptable. Se evidenció un valor de R^2 promedio de 0.8, el cual indica que una proporción de la variabilidad de las concentraciones reales de los tres analitos puede ser explicada por el modelo. Aunque, lo óptimo es que el coeficiente de determinación sea mayor al 0.9, lo cual es consistente con lo descrito por Duarte et al. (2002), quienes obtuvieron métricas superiores a 0.95 en la cuantificación de azúcares en jugos mediante FTIR-ATR. Paralelamente, los resultados del RMSE indican que hay una desviación entre los valores observados y predichos, pero resulta complejo alcanzar un RMSE perfecto en un amplio rango de valores numéricos.

En las gráficas de observado versus predicho, los puntos tienden a distribuirse alrededor de la línea de identidad, lo cual indica que el modelo presenta una adecuada capacidad para capturar la tendencia global de la relación entre la señal espectral y la concentración de los analitos. Para el caso de la sacarosa, la dispersión de los puntos es menor y se aprecia una alineación más cercana a la recta ideal, lo que sugiere un mejor ajuste del modelo para este azúcar. En contraste, las gráficas correspondientes a glucosa y fructosa muestran una mayor dispersión, lo cual indica una menor exactitud relativa en la predicción individual de estos analitos. Este comportamiento puede atribuirse al solapamiento espectral entre azúcares y a su menor contribución relativa dentro de las mezclas, limitando la capacidad del modelo para discriminar de forma precisa entre compuestos con estructuras químicas similares.

En la validación externa de los modelos con muestras de bebidas comerciales, las cuales cabe recalcar son matrices mucho más complejas que pueden presentar una concentración mayor a la recomendada por los organismos regulatorios, se presentó una tendencia en la que el modelo arrojaba valores negativos en la concentración de los analitos, acentuándose en la sacarosa. Aquello no implica que los azúcares no estén presentes en las bebidas, sino que los coeficientes de regresión pueden producir predicciones negativas cuando las muestras están fuera de las condiciones de calibración y es complejo interpretar vectores de regresión en métodos de calibración inversa a menos de que exista la suficiente información (Brown & Green, 2009). Entonces, el modelo no está diseñado para ser capaz de extrapolar valores fuera del rango de calibración, lo cual conlleva a obtener predicciones erráticas o pobres (Olivieri, 2021).

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Emplear el rango espectral correspondiente a la huella digital de los carbohidratos para el desarrollo del modelo contribuyó a reducir el ruido, mejorar la relación entre las variables espectrales y la de respuesta y, en general, a obtener una mejor precisión.
- A pesar de que el modelo mostró valores aceptables de error cuadrático medio y coeficiente de determinación durante la fase de prueba, en la validación el modelo tuvo un desempeño menor, aludiendo a que la capacidad predictiva no es precisa para muestras con concentración variable a la establecida en el diseño de mezclas.
- El modelo se vio limitado, principalmente, por la poca variabilidad del conjunto de datos destinados al entrenamiento, las similitudes estructurales de los analitos y el reducido número de datos espectrales proporcionados.
- Los resultados indican que no resulta viable crear modelos de regresión multidimensional que predigan simultáneamente glucosa, fructosa y sacarosa en bebidas azucaradas, a menos que se optimice el flujo de trabajo y que se proporcionen más datos espectrales a diversas concentraciones.

5.2. Recomendaciones

- Incrementar el número de muestras para el entrenamiento y prueba del modelo, de forma que exista más información espectral con la que el modelo pueda aprender y, de este modo, optimizar la predicción y mejorar los valores del RMSE y R^2 .

- Reentrenar el modelo con más patrones a diversas concentraciones superiores al 6%, para que aprenda más relaciones espectrales y pueda predecir de mejor forma las bebidas comerciales.

DECLARACIÓN Y DISPONIBILIDAD DEL CÓDIGO

El código empleado para el desarrollo de la presente investigación se encuentra disponible en el siguiente repositorio:

<https://github.com/alligr17/Tesis>

REFERENCIAS

- Ando, Y., Ohta, Y., Munetsuna, E., Yamada, H., Nouchi, Y., Kagemaya, I., Mizuno, G., Yamazai, M., Fujii, R., Ishikawa, H., Suzuki, K., Hashimoto, S., & Ohashi, K. (2023). Laboratory analysis of glucose, fructose, and sucrose contents in Japanese common beverages for the exact assessment of beverage-derived sugar intake. *Fujita Medical Journal*, 9(2), 126–133.
- Arias, F. (2012). *El proyecto de investigación. Introducción a la metodología científica* (6th ed.). Editorial Episteme.
- Assis, C., Pereira, H. V., Amador, V. S., Augusti, R., de Oliveira, L. S., & Sena, M. M. (2019). Combining mid infrared spectroscopy and paper spray mass spectrometry in a data fusion model to predict the composition of coffee blends. *Food Chemistry*, 281, 71–77. <https://doi.org/10.1016/j.foodchem.2018.12.044>
- Baek, S.-J., Park, A., Ahn, Y.-J., & Choo, J. (2015). Baseline correction using asymmetrically reweighted penalized least squares smoothing. *The Analyst*, 140(1), 250–257. <https://doi.org/10.1039/C4AN01061B>
- Becerra, M., & Goos, P. (2021). Bayesian I-optimal designs for choice experiments with mixtures. *Chemometrics and Intelligent Laboratory Systems*, 217, 104395. <https://doi.org/10.1016/j.chemolab.2021.104395>
- Bolaños-Quirós, F., Fallas-Rodríguez, P., Murillo-González, L., Pérez, A. M., & Quirós, A. M. (2024). Desarrollo de una bebida verde con espirulina (*Arthrospira* sp.) mediante el diseño experimental de mezclas. *Agronomía Mesoamericana*, 60115. <https://doi.org/10.15517/am.2024.60115>

- Brown, C. D., & Green, R. L. (2009). Critical factors limiting the interpretation of regression vectors in multivariate calibration. *TrAC Trends in Analytical Chemistry*, 28(4), 506–514. <https://doi.org/10.1016/j.trac.2009.02.003>
- Cabello-Solorzano, K., Ortigosa de Araujo, I., Peña, M., Correia, L., & J. Tallón-Ballesteros, A. (2023). *The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis* (pp. 344–353). https://doi.org/10.1007/978-3-031-42536-3_33
- Calle, J. L. P., Ferreiro-González, M., Ruiz-Rodríguez, A., Fernández, D., & Palma, M. (2022). Detection of Adulterations in Fruit Juices Using Machine Learning Methods over FT-IR Spectroscopic Data. *Agronomy*, 12(3), 683. <https://doi.org/10.3390/agronomy12030683>
- Chavarro, B., & Vargas, A. (2022). Evaluar el Software Libre R para el Análisis de Espectros Infrarrojo Empleando Quimiometría. *Revista Ingenierías USBMed*, 13(2), 48–55.
- Christian, G., Dasgupta, P., & Schug, K. (2014). *Analytical Chemistry* (7th ed.). John Wiley & Sons, Inc.
- Christin Brettschneider, K., Zettel, V., Sadeghi Vasafi, P., Hummel, D., Hinrichs, J., & Hitzmann, B. (2022). Spectroscopic-Based Prediction of Milk Foam Properties for Barista Applications. *Food and Bioprocess Technology*, 15(8), 1748–1757. <https://doi.org/10.1007/s11947-022-02822-3>
- Coldea, T. E., Socaciu, C., Fetea, F., Ranga, F., Pop, R. M., & Florea, M. (2013). Rapid Quantitative Analysis of Ethanol and Prediction of Methanol Content in Traditional Fruit Brandies from Romania, using FTIR Spectroscopy and Chemometrics.

Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 41(1), 143.

<https://doi.org/10.15835/nbha4119000>

Cornell, J. (2011). *A Primer on Experiments with Mixtures*. John Wiley & Sons, Inc.

Damto, T., Zewdu, A., & Birhanu, T. (2023). Application of Fourier transform infrared (FT-IR) spectroscopy and multivariate analysis for detection of adulteration in honey markets in Ethiopia. *Current Research in Food Science*, 7, 100565.

<https://doi.org/10.1016/j.crfs.2023.100565>

Davies, A. M. C. (2007). Back to basics: spectral pre-treatments, derivatives. *Tony*

Davies Column, 19(2), 32–33.

de Oliveira, B. P., Lelis, E., & Schons, E. (2026). Prediction of Li₂O and Spodumene by FTIR-PLS in Pegmatitic Samples for Process Control. *Minerals*, 16(1), 66.

<https://doi.org/10.3390/min16010066>

Delwiche, S., & Reeves, J. (2010). A Graphical Method to Evaluate Spectral Preprocessing in Multivariate Regression Calibrations: Example with Savitzky–Golay Filters and Partial Least Squares Regression. *Applied Spectroscopy*, 64(1), 73–82.

Dhaulaniya, A. S., Balan, B., Yadav, A., Jamwal, R., Kelly, S., Cannavan, A., & Singh, D. K. (2020). Development of an FTIR based chemometric model for the qualitative and quantitative evaluation of cane sugar as an added sugar adulterant in apple fruit juices. *Food Additives & Contaminants: Part A*, 37(4), 539–551.

<https://doi.org/10.1080/19440049.2020.1718774>

Duarte, I. F., Barros, A., Delgadillo, I., Almeida, C., & Gil, A. M. (2002). Application of FTIR Spectroscopy for the Quantification of Sugars in Mango Juice as a Function

of Ripening. *Journal of Agricultural and Food Chemistry*, 50(11), 3104–3111.

<https://doi.org/10.1021/jf011575y>

Frost, T. (1999). Quantitative Analysis. In *Encyclopedia of Spectroscopy and Spectrometry* (pp. 1931–1936). Elsevier. <https://doi.org/10.1006/rwsp.2000.0250>

Greenacre, M., Groenen, P. J. F., Hastie, T., D’Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1), 100. <https://doi.org/10.1038/s43586-022-00184-w>

Guerrero Niño, J. L., Garro Duarte, Ó. A., Caballero Blanco, K., Medina Barragán, L. M., & Garcia, J. A. (2025). Optimización del control de procesos en la agroindustria mediante espectroscopia FTIR. *Palmas*, 45(4), 39–43. <https://doi.org/10.56866/01212923.14426>

Huang, Y. (2022). Chemometric Methods in Analytical Spectroscopy Technology. In *Chemometric Methods in Analytical Spectroscopy Technology* (pp. 1–29). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-1625-0_1

Jiménez, H., Carpintero, E. de J., Campos, V. Y., Gomez, R. E., & Ayala, A. (2022). Predicción de la coloración de la granola, mediante un modelo multivariado para la industria de alimentos. *Ciencia Latina Revista Científica Multidisciplinar*, 6(5), 5412–5428. https://doi.org/10.37811/cl_rcm.v6i5.3512

Koirala, R. P., Dawanse, S., & Pantha, N. (2022). Diffusion of glucose in water: A molecular dynamics study. *Journal of Molecular Liquids*, 345, 117826. <https://doi.org/10.1016/j.molliq.2021.117826>

- Krishnan, S. R., & Seelamantula, C. S. (2013). On the Selection of Optimum Savitzky-Golay Filters. *IEEE Transactions on Signal Processing*, *61*(2), 380–391.
<https://doi.org/10.1109/TSP.2012.2225055>
- Lara-Cruz, G. A., & Jaramillo-Botero, A. (2022). Molecular Level Sucrose Quantification: A Critical Review. *Sensors*, *22*(23), 9511.
<https://doi.org/10.3390/s22239511>
- Lara-Serrano, M., Morales-delaRosa, S., Campos-Martin, J. M., Abdelkader-Fernández, V. K., Cunha-Silva, L., & Balula, S. S. (2021). Isomerization of glucose to fructose catalyzed by metal–organic frameworks. *Sustainable Energy & Fuels*, *5*(15), 3847–3857. <https://doi.org/10.1039/D1SE00771H>
- Lê Cao, K.-A., Boitard, S., & Besse, P. (2011). Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*, *12*(1), 253. <https://doi.org/10.1186/1471-2105-12-253>
- Leopold, L., Leopold, N., Diehl, H.-A., & Socaciu, C. (2011). Quantification of carbohydrates in fruit juices using FTIR spectroscopy and multivariate analysis. *Journal of Spectroscopy*, *26*(2), 93–104.
- Matienzo, J. A. (2021). Method based on Raman spectroscopy and partial least squares regression to predict the authenticity of bee honey. *Agroindustrial Science*, *11*(2), 169–173. <https://doi.org/10.17268/agroind.sci.2021.02.05>
- Mayorga, T., Lascano, N., Valencia, A., & Robalino, D. (2022). Tendencia del consumo de las bebidas azucaradas en el Ecuador 2014-2019. *Revista Uniandes Episteme*, *9*(4), 589–601.

- Mohamed, M. A., Jaafar, J., Ismail, A. F., Othman, M. H. D., & Rahman, M. A. (2017). Fourier Transform Infrared (FTIR) Spectroscopy. In *Membrane Characterization* (pp. 3–29). Elsevier. <https://doi.org/10.1016/B978-0-444-63776-5.00001-2>
- Mondragón, P. (2020). *La espectroscopia FTIR-ATR aplicada al análisis de alimentos y bebidas*.
- Moros, J. (2007). *TRATAMIENTO NUMÉRICO DE LOS DATOS EN EL ANÁLISIS CUANTITATIVO POR ESPECTROMETRÍA VIBRACIONAL* [Tesis Doctoral]. Universitat de València.
- OECD-FAO. (2009). OECD-FAO Agricultural Outlook. In *OECD Agriculture Statistics*. OECD Publishing. <https://doi.org/10.1787/agr-outl-data-en>
- Olivieri, A. (2021). *Calibraciones univariada y multivariada de primer orden*. Universidad Nacional de Rosario.
- Orozco-Morales, R., Ojito-Ramos, K., Aday, O., & Santana, L. H. (2025). Caracterización espectral de arvenses que afectan los cultivos de *Saccharum spp.*, métodos de obtención y procesamiento de datos: Una revisión. *Centro Agrícola*, 52, e2458.
- Peng, J., Peng, S., Jiang, A., Wei, J., Li, C., & Tan, J. (2010). Asymmetric least squares for multiple spectra baseline correction. *Analytica Chimica Acta*, 683(1), 63–68. <https://doi.org/10.1016/j.aca.2010.08.033>
- Pérez-López, E. (2023). Un método cromatográfico para la determinación de sacarosa en bebidas comerciales con azúcar añadida. *Pensamiento Actual*, 23(40).

- Prada, J. (2015). *ANÁLISIS METABOLÓMICO DE LA ESPECIE Baccharis latifolia (ASTERACEAE) EN LA SABANA DE BOGOTÁ* [Trabajo de Grado]. Universidad Militar Nueva Granada.
- Roman, V. (2019). *Machine Learning Supervisado: Fundamentos de la Regresión Lineal*. <https://medium.com/datos-y-ciencia/machine-learning-supervisado-fundamentos-de-la-regresi%C3%B3n-lineal-bbcb07fe7fd>.
- Rosales-Martínez, O., Granda-Gutiérrez, E. E., García-Hernández, R. A., Alejo-Eleuterio, R., & Flores-Fuentes, A. A. (2025). Spectral Derivatives Improve FTIR-Based Machine Learning Classification of Plastic Polymers. *Modelling*, 6(4), 115. <https://doi.org/10.3390/modelling6040115>
- Sahlan, M., Karwita, S., Gozan, M., Hermansyah, H., Yohda, M., Yoo, Y. J., & Pratami, D. K. (2019). Identification and classification of honey's authenticity by attenuated total reflectance Fourier-transform infrared spectroscopy and chemometric method. *Veterinary World*, 12(8), 1304–1310. <https://doi.org/10.14202/vetworld.2019.1304-1310>
- Savitzky, Abraham., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Shen, X., Ye, S., Xu, L., Hu, R., Jin, L., Xu, H., Liu, J., & Liu, W. (2018). Study on baseline correction methods for the Fourier transform infrared spectra with different signal-to-noise ratios. *Applied Optics*, 57(20), 5794. <https://doi.org/10.1364/AO.57.005794>
- Skoog, D., West, D., Holler, F., & Crouch, S. (2022). *Fundamentals of Analytical Chemistry* (9th ed.). Cengage Learning.

- Stuart, B. (2021). Infrared Spectroscopy. In *Analytical Techniques in Forensic Science* (pp. 145–160). Wiley. <https://doi.org/10.1002/9781119373421.ch7>
- Szaniawska, A., Grzeda, J., Binder, J., Kudelski, A., Malek, K., Wrobel, T. P., Wysmolek, A., & Roguz, K. (2025). Data-Driven FTIR Spectroscopy for the Discrimination of Nectars. *Molecules*, *30*(20), 4083. <https://doi.org/10.3390/molecules30204083>
- Szymanska-Chargot, M., & Zdunek, A. (2013). Use of FT-IR Spectra and PCA to the Bulk Characterization of Cell Wall Residues of Fruits and Vegetables Along a Fraction Process. *Food Biophysics*, *8*(1), 29–42. <https://doi.org/10.1007/s11483-012-9279-7>
- UNICEF Ecuador. (2024). *Política pública enfocada en la promoción de entornos alimentarios saludables para la prevención del sobrepeso y obesidad en niñas, niños y adolescentes en Ecuador*. UNICEF Ecuador .
- Workman, Jr. , J., & Weyer, L. (2007). *Practical Guide to Interpretive Near-Infrared Spectroscopy*. CRC Press. <https://doi.org/10.1201/9781420018318>
- World Bank. (2020). *Taxes on Sugar-Sweetened Beverages: International Evidence and Experiences*. World Bank.
- World Health Organization. (2015). *Guideline: Sugars Intake for Adults and Children. Executive summary*. World Health Organization.
- Worley, B., & Powers, R. (2012). Multivariate Analysis in Metabolomics. *Current Metabolomics*, *1*(1), 92–107. <https://doi.org/10.2174/2213235X11301010092>

Zhang, D. (2017). A Coefficient of Determination for Generalized Linear Models. *The American Statistician*, 71(4), 310–316.

<https://doi.org/10.1080/00031305.2016.1256839>

Zhang, Z.-M., Chen, S., & Liang, Y.-Z. (2010). Baseline correction using adaptive iteratively reweighted penalized least squares. *The Analyst*, 135(5), 1138.

<https://doi.org/10.1039/b922045c>

ANEXOS

Anexo 1

Informe de resultados del análisis cromatográfico de una de las muestras comerciales.



INFORME DE RESULTADOS

INF.AFQ.29151a

DATOS DEL CLIENTE

Cientes:	ROSA PESANTES
Dirección:	AV. TURUHUAICO Y CALLE VIEJA, CUENCA.
Teléfono:	0987525111

DATOS DE LA MUESTRA

Nombre de la Muestra:	COCA COLA SABOR ORIGINAL	Lote:	L2CN
Tipo de muestra:	COCA COLA	Fecha elaboración:	18ENE26
Muestreado por:	CLIENTE	Fecha vencimiento:	X
Color:	CARACTERISTICO	Contenido declarado:	200 ml
Olor:	CARACTERISTICO	Contenido encontrado:	200 ml
Estado:	LIQUIDO	Fecha de recepción:	2026-01-26
		Hora de recepción:	13:23:38
		Fecha análisis:	2026-01-27
		Fecha entrega:	2026-01-29

RESULTADOS FISICOQUIMICOS

PARAMETRO	RESULTADO	UNIDAD	METODO DE ANALISIS INTERNO	METODO DE ANALISIS DE REFERENCIA	INCERTIDUMBRE
AZUCARES TOTALES	9,80	%	PA-FQ-39	AOAC 977.20 Modificado	----
LACTOSA	0,00	%	PA-FQ-39	AOAC 977.20 Modificado	----
SACAROSA*	8,78	%	PA-FQ-39	AOAC 977.20 Modificado	+/-1,12
FRUCTOSA*	1,02	%	PA-FQ-39	AOAC 977.20 Modificado	+/-0,14
GLUCOSA	0,00	%	PA-FQ-39	AOAC 977.20 Modificado	----

Nota 1: La información de datos del cliente y de la muestra que afecte a la validez de resultados es proporcionada y exclusiva responsabilidad del cliente y no representa responsabilidad para EcuachemLab Cia. Ltda.

Nota 2: Sin la aprobación escrita del Laboratorio no se debe reproducir el informe, excepto cuando se reproducen en su totalidad.

Nota 3: Los ensayos marcados con (*) NO están incluidos en el alcance de la acreditación del SAE.

Nota 4: El resultado se refiere únicamente a la muestra recibida o tomada por laboratorio, Ecuachemlab Cia. Ltda., se responsabiliza exclusivamente de los análisis

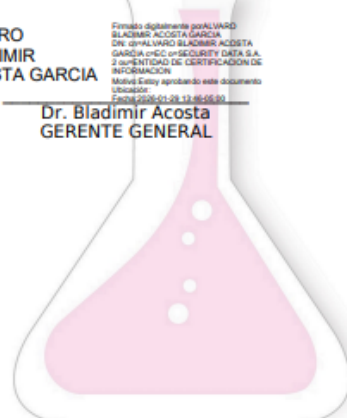
JENNY
ADRIANA
ROBALINO
PUENTE

Quim. Alim. Jenny Robalino
JEFE SECCION INSTRUMENTAL

ALVARO
BLADIMIR
ACOSTA GARCIA

Dr. Bladimir Acosta
GERENTE GENERAL

Formado digitalmente por ALVARO
BLADIMIR ACOSTA GARCIA
DN: cn=ALVARO BLADIMIR ACOSTA
GARCIA, o=CC SECURITY DATA S.A.
2 AUTENTICIDAD DE CERTIFICACION DE
INFORMACION
Motivo: Estoy aprobando este documento
Usuario:
Fecha: 2026-01-28 13:46:06.00



Síguenos:   