



UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE BIOTECNOLOGÍA

**“MODELAMIENTO PREDICTIVO MULTIDIMENSIONAL PARA LA
CUANTIFICACIÓN SIMULTÁNEA DE ÁCIDO ACÉTICO Y ETANOL EN
VINAGRE DE MANZANA MEDIANTE INFRARROJOS POR TRANSFORMADA
DE FOURIER”**

Trabajo de titulación previo a la obtención

del título de Ingeniera Biotecnóloga

AUTORA: GABRIELA LISSETH GARCÍA LOJA

TUTOR: ING. PABLO WILSON ARÉVALO MOSCOSO, PhD.

Cuenca - Ecuador

2025

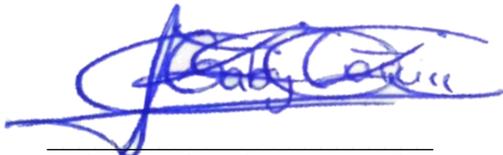
**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, Gabriela Lisseth García Loja con documento de identificación N° 0707073607
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, 19 de febrero del 2025

Atentamente,



Gabriela Lisseth García Loja

0707073607

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

Yo, Gabriela Lisseth García Loja con documento de identificación N° 0707073607, 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: “Modelamiento predictivo multidimensional para la cuantificación simultánea de ácido acético y etanol en vinagre de manzana mediante infrarrojos por transformada de Fourier”, 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, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 19 de febrero del 2025

Atentamente,



Gabriela Lisseth García Loja

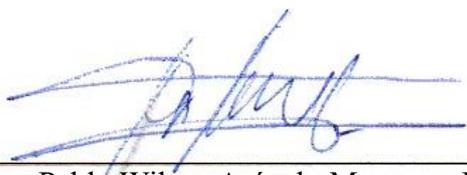
0707073607

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: “MODELAMIENTO PREDICTIVO MULTIDIMENSIONAL PARA LA CUANTIFICACIÓN SIMULTÁNEA DE ÁCIDO ACÉTICO Y ETANOL EN VINAGRE DE MANZANA MEDIANTE INFRARROJOS POR TRANSFORMADA DE FOURIER”, realizado por Gabriela Lisseth García Loja con documento de identificación N° 0707073607, obteniendo como resultado final el trabajo de titulación bajo la opción Trabajo experimental que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 19 de febrero del 2025

Atentamente,



Ing. Pablo Wilson Arévalo Moscoso, PhD.

0102156957

DEDICATORIA

A Dios, por permitirme alcanzar una meta más, por brindarme esa sabiduría e inteligencia, por guiarme y protegerme todo este tiempo lejos de casa.

A mis padres, que nunca dudaron en mi potencial, por brindarme la fuerza necesaria y ánimo cuando ya estaba por rendirme, por ese amor incondicional y consejos que nunca faltaron.

A mi hermano quien sin duda ha sido mi cómplice y compañero en esta vida de ‘foráneo’, mi más sincera gratitud ante tu ayuda, aunque a veces me sacas de quicio, agradezco tus pequeños detalles, ocurrencias y compañía.

A mi familia, García y Loja, quienes siempre me recibían con los brazos abiertos cada que regresaba a casa, en especial a mi prima Judith, quien siempre estaba atenta a escuchar cada una de mis aventuras.

A mi segunda familia, la que formé en mi tiempo académico, conformada por las personas que han marcado en mi vida y me han acompañado en cada una de mis aventuras, momentos de tristeza y alegría; ellos hicieron ameno mi estadía en una ciudad distinta y en todo este tiempo de formación académica y personal, gracias a: Diego, Jasson, Edmundo, Angy, Sofy, Milena, Gustavo, Accel, Leslie, Carlos, Paúl, Dailenis, Wendy, William, Josselyn. Gracias a cada uno de ustedes por su Respaldo incondicional y su Entrega sincera, porque en algún momento me alegraron el día con alguna ocurrencia o con su compañía; sea donde sea mi próxima aventura, siempre los llevaré en mi corazón.

A mis amistades de Machala: Diana L., Itaty L., Romina V. y Roberto C., con quienes a lo mejor no hablo todos los días, pero siempre siento su apoyo desde la distancia.

A mis compañeros y amigos de la carrera y trabajo, que, aunque no fueron nombrados, agradezco su apoyo y presencia en este momento importante.

A mis profesores, quienes fueron guías fundamentales para la realización de este trabajo. Un agradecimiento especial a mi profesor Ing. Edmond Geraud, quien con mucha paciencia me guió en este trabajo.

AGRADECIMIENTOS

Deseo expresar mi más profundo agradecimiento a Dios y a todas las personas que me acompañaron y apoyaron en la realización de este proyecto de tesis, que marca no solo el cumplimiento de una meta académica, sino también el cierre de una etapa significativa de mi vida.

En primer lugar, quiero agradecer al Ing. Edmond Geraud, cuya guía, paciencia y conocimientos fueron fundamentales para el desarrollo y la culminación de este trabajo. Su apoyo incondicional y sus valiosos consejos fueron clave en cada paso del proceso.

También agradezco a mi tutor de tesis, Ing. Pablo Arévalo, por sus observaciones y aportaciones que enriquecieron enormemente mi trabajo. Su disposición ha sido de gran ayuda para alcanzar los estándares requeridos.

Un agradecimiento especial a mis compañeros de laboratorio y amigos, quienes compartieron conmigo muchas horas de trabajo y desafíos. Gracias por su amistad, compañerismo y todos los momentos de alegría y estrés que hemos pasado.

Para aquellas personas que en su momento me acompañaron y por alguna razón ya no se encuentran cerca mío, les agradezco de corazón haber estado presente en mi vida y enseñarme muchas cosas. Cada una de esas experiencias compartidas ha dejado una huella imborrable en mi persona. (W.G y R.E, muchas gracias)

No puedo dejar de mencionar a mi familia, mis padres (Gabriel y Sara), mis abuelitos (Domitila y Ramón), mi hermano David, tías/os y primos en general, que han sido mi soporte emocional y motivación constante.

Finalmente, agradezco a todas las personas que directa o indirectamente contribuyeron a este proyecto. Cada uno de ustedes ha sido parte esencial de este viaje.

Tabla de contenido

ABREVIATURAS.....	10
RESUMEN.....	11
ABSTRACT.....	11
CAPÍTULO 1.....	12
INTRODUCCIÓN.....	12
1.1 PLANTEAMIENTO DEL PROBLEMA.....	12
1.3 PREGUNTA DE INVESTIGACIÓN.....	13
1.4 JUSTIFICACIÓN.....	13
1.5 LIMITACIONES.....	14
1.6 OBJETIVOS.....	15
1.6.1 GENERAL.....	15
1.6.2 ESPECÍFICOS.....	15
1.7 HIPÓTESIS.....	15
CAPÍTULO 2.....	16
FUNDAMENTACIÓN TEÓRICA.....	16
2.1. ESTADO DEL ARTE.....	16
2.2 MARCO TEÓRICO.....	18
2.3 MARCO CONCEPTUAL.....	28
CAPÍTULO 3.....	29
MATERIALES Y MÉTODOS.....	29
3.1 NIVEL DE INVESTIGACIÓN.....	29
3.2 DISEÑO DE INVESTIGACIÓN.....	29
3.3 DISEÑO DE EXPERIMENTO.....	29
3.4 VARIABLES.....	32
3.4.1 VARIABLES INDEPENDIENTES.....	32

3.4.2 VARIABLES DEPENDIENTES	32
3.4.3 VARIABLES INTERVINIENTES	33
3.4.4 VARIABLES EXTRAÑAS	33
3.5 TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS	33
3.6 TÉCNICAS DE PROCESAMIENTO Y ANÁLISIS DE DATOS	33
3.7 PROTOCOLO A IMPLEMENTAR	34
3.7.1 FASE PREEXPERIMENTAL	34
3.7.2 FASE EXPERIMENTAL.....	35
3.7.3 MÉTODOS ESTADÍSTICOS	40
CAPÍTULO 4.....	46
RESULTADOS Y DISCUSIÓN	46
4.1 RESULTADOS	46
4.1.1 SELECCIÓN DE RANGOS ÓPTIMOS	46
4.1.2 ANÁLISIS PREEXPLORATORIO.....	47
4.1.3 RESULTADOS DEL ANÁLISIS PLS-DA	50
4.1.4 RESULTADOS DEL ANÁLISIS sPLS	52
4.1.5 PREDICCIÓN DE MUESTRAS DE VINAGRE	55
4.2 DISCUSIÓN	57
4.2.1 DETERMINACIÓN DE RANGOS ESPECTRALES ÓPTIMOS....	57
4.2.2 INTERPRETACIÓN DE MODELOS PREDICTIVOS MULTIDIMENSIONALES	58
4.2.3 EVALUACIÓN DEL MODELO PLS Y PREDICCIÓN EN MUESTRAS DE VINAGRE.....	59
4.2.4 COMPARACIÓN CON ESTUDIOS PREVIOS Y LIMITACIONES DEL MODELO	59
CAPÍTULO 5.....	60
5.1 CONCLUSIONES.....	60
5.2 RECOMENDACIONES.....	61

DECLARACIÓN DE DISPONIBILIDAD DEL CÓDIGO	62
REFERENCIAS BIBLIOGRÁFICAS	63
ANEXOS	67
ANEXO 1:	67
ANEXO 2:	68

ABREVIATURAS

- **ALS:** Suavizado por Mínimos Cuadrados Alternados (Asymmetric Least Squares Smoothing)
- **CODEX:** Codex Alimentarius
- **FTIR:** Espectroscopía Infrarroja por Transformada de Fourier (Fourier Transform Infrared Spectroscopy)
- **INEN:** Instituto Ecuatoriano de Normalización
- **PCA:** Análisis de Componentes Principales (Principal Component Analysis)
- **PLS:** Regresión por Mínimos Cuadrados Parciales (Partial Least Squares Regression)
- **PLS-DA:** Análisis Discriminante por Mínimos Cuadrados Parciales (Partial Least Squares Discriminant Analysis)
- **ROC:** Característica Operativa del Receptor (Receiver Operating Characteristic)
- **SG:** Filtro de Savitzky–Golay
- **sPLS:** Regresión por Mínimos Cuadrados Parciales Escasa (Sparse Partial Least Squares Regression)
- **sPLS-DA:** Análisis Discriminante por Mínimos Cuadrados Parciales Escaso (Sparse Partial Least Squares Discriminant Analysis)

RESUMEN

La industria alimentaria gasta una cantidad considerable de dinero en control de calidad para garantizar el contenido y la seguridad del producto. Este podría ser el caso del vinagre, donde elementos como contaminantes de etanol pueden afectar su clasificación y comercialización.

Para la detección y cuantificación de etanol en vinagre, se utilizaron técnicas de espectroscopía FTIR y modelado matemático. Se construyeron modelos predictivos de sPLS y PLS-DA mientras se identificaban los mejores métodos de preprocesamiento espectral (ALS, normalización, suavizado y segunda derivada de Savitzky-Golay).

Los modelos se evaluaron utilizando métricas de evaluación como RMSE y curvas ROC que resultaron en una separación efectiva entre muestras contaminadas y no contaminadas. Los hallazgos muestran que FTIR con análisis multivariante es un método rápido y preciso que puede emplearse para detectar etanol, útil para ofrecer una solución de bajo costo en contraste con métodos más laboriosos.

ABSTRACT

The food industry spends a considerable amount of money on quality control to ensure product content and safety. This could be the case for vinegar, where elements such as ethanol contaminants can affect its classification and marketing.

For the detection and quantification of ethanol in vinegar, FTIR spectroscopy techniques and mathematical modeling were used. Predictive sPLS and PLS-DA models were built while the best spectral preprocessing methods (ALS, normalization, smoothing, and Savitzky-Golay second derivative) were identified.

The models were evaluated using evaluation metrics such as RMSE and ROC curves that resulted in an effective separation between contaminated and uncontaminated samples. The findings show that FTIR with multivariate analysis is a fast and accurate method that can be employed to detect ethanol, useful in offering a low-cost solution in contrast to more laborious methods.

CAPÍTULO 1

INTRODUCCIÓN

1.1 PLANTEAMIENTO DEL PROBLEMA

En la industria alimentaria, el control y la mejora de los procesos de producción son esenciales para garantizar la calidad del producto final. Es crucial supervisar la seguridad de los alimentos y asegurar que no representen riesgos para la salud del consumidor. Para lograrlo, los sistemas de control de calidad deben ser rigurosos y sistemáticos. Sin embargo, los métodos tradicionales utilizados para este fin, como la cromatografía o las titulaciones, suelen ser costosos y requieren un tiempo considerable para obtener resultados, lo que genera la necesidad de implementar métodos más rápidos y económicos.

La producción de vinagre no es una excepción a estos desafíos. El vinagre es sometido a procesos de fermentación controlada, durante los cuales es fundamental monitorear parámetros clave para asegurar la calidad del producto. Si el proceso de fermentación no se regula adecuadamente, pueden aparecer contaminantes como metanol, etanol (alcoholes residuales), arsénico o plomo, los cuales comprometen la seguridad del producto. (Codex Alimentarius, 1987) Las normativas internacionales, como el Codex Alimentarius Stan 162-1987, y las nacionales, como las normas INEN 2296:2013, establecen límites estrictos sobre estos contaminantes. En particular, el etanol debe mantenerse por debajo de un 0,5 % en vinagres comerciales y un máximo de 1 % en vinagres de vino (INEN, 2013).

Esta inclusión no controlada puede alterar la clasificación del producto. Si el contenido de etanol es significativo, el vinagre podría considerarse una bebida alcohólica en lugar de un simple condimento, lo que conllevaría a regulaciones diferentes en términos de etiquetado y comercialización. Esto podría complicar la distribución y requerir licencias específicas para su venta.

1.3 PREGUNTA DE INVESTIGACIÓN

¿Existen relaciones funcionales entre las longitudes de onda y las concentraciones de ácido acético y etanol que permitan generar modelos de regresión para la predicción simultánea de sus límites legales en vinagre comercial?

1.4 JUSTIFICACIÓN

La calidad y seguridad de los productos alimenticios son aspectos fundamentales en la industria, ya que cualquier desviación en los parámetros establecidos puede comprometer tanto la salud del consumidor como la reputación de las empresas productoras. En el caso del vinagre, es esencial controlar su proceso de producción para evitar la presencia de contaminantes como alcoholes o metales pesados. Las normativas nacionales e internacionales, como las establecidas por el Codex Alimentarius y las normas INEN 2296:2013, exigen que los niveles de etanol no excedan el 0,5 % en vinagres comerciales y el 1 % en vinagres de vino, lo que subraya la importancia de un monitoreo preciso y constante. (Codex Alimentarius, 1987) (INEN, 2013)

Los métodos convencionales de análisis, como la cromatografía y la titulación, si bien efectivos, son costosos, requieren tiempo y demandan personal especializado, lo que puede limitar su aplicación para un control de calidad continuo y en tiempo real, por lo que la espectroscopía infrarroja con transformada de Fourier (FTIR) se presenta como una alternativa viable debido a su capacidad para realizar análisis rápidos, no destructivos y precisos. (Movasaghi et al., 2008) Sin embargo, el desarrollo de modelos predictivos que puedan correlacionar las firmas espectrales con las concentraciones de analitos aún presenta un desafío en el campo de la industria alimentaria.

Este proyecto es relevante porque busca optimizar el uso de la FTIR para la detección de contaminantes, particularmente etanol, en el vinagre de manzana. La

implementación de un modelo matemático que permita predecir de manera eficiente las concentraciones de etanol a partir de los espectros infrarrojos reduciría significativamente los costos y tiempos de análisis, lo que facilitaría el control de calidad en tiempo real. Esto no solo beneficiaría a las empresas productoras de vinagre, mejorando la eficiencia de sus procesos y asegurando el cumplimiento de las normativas vigentes, sino que también garantizaría un producto más seguro para los consumidores.

Además, los resultados de esta investigación podrían ser extrapolables a otras industrias donde se requiera un análisis rápido y preciso de compuestos en soluciones líquidas, abriendo la puerta a futuras aplicaciones en diferentes sectores de la industria alimentaria y química. (Berthomieu & Hienerwadel, 2009) Por tanto, este estudio no solo contribuirá al mejoramiento del proceso productivo del vinagre, sino que también ofrecerá nuevas herramientas tecnológicas para optimizar los controles de calidad en otras áreas.

1.5 LIMITACIONES

El estudio actual tiene varias limitaciones notables, una de las principales desventajas sería la accesibilidad a los instrumentos apropiados necesarios para llevar a cabo el análisis de las muestras, lo que limitaría el número de análisis o la velocidad a la que se pueden realizar.

Además, la detección de algunos analitos objetivo, incluido el etanol y el ácido acético, puede estar limitada por interferencias espectrales de otros componentes del vinagre. Estas interferencias pueden hacer que el modelo predictivo sea inexacto y no podamos confiar en sus resultados.

Otra limitación crucial es el tiempo para llevar a cabo la investigación. En caso de que el tiempo asignado no sea suficiente para completar fases como el análisis de muestras, construcción del modelo y validación, la profundidad del estudio se verá comprometida y el número de muestras procesadas disminuirá, lo que afectará la calidad de la investigación y los resultados finales.

1.6 OBJETIVOS

1.6.1 GENERAL

Establecer relaciones funcionales entre las longitudes de onda y las concentraciones de ácido acético y etanol en los espectros de estándares de vinagre, mediante la implementación de modelos predictivos multidimensionales para su cuantificación simultánea en muestras.

1.6.2 ESPECÍFICOS

- Determinar los rangos espectrales óptimos mediante el análisis y diferenciación de las bandas características de los analitos, para optimizarlos con tratamiento matemático.
- Formular modelos predictivos multidimensionales a partir de los datos seleccionados y optimizados, utilizando plataformas de desarrollo para realizar predicciones simultáneas.
- Evaluar la precisión predictiva de los modelos de regresión utilizando muestras propias de vinagre (ácido acético) para autenticar la exactitud de los modelos generados.

1.7 HIPÓTESIS

Hay una relación funcional entre las longitudes de onda y las concentraciones de diferentes analitos objetivo, lo que permite desarrollar modelos de predicción multidimensionales.

CAPÍTULO 2

FUNDAMENTACIÓN TEÓRICA

2.1. ESTADO DEL ARTE

La aplicación del FTIR es muy amplia y versátil, ya que es una técnica de gran eficiencia, simple y precisa que es utilizada para la caracterización de compuestos químicos en diversas matrices. Su principio fundamental se basa en la absorción de radiación infrarroja por las moléculas, lo que genera bandas espectrales características que permiten identificar grupos funcionales presentes en una muestra (Smok-Kalwat et al., 2024).

En este estudio, se aplicó el FTIR para el análisis de muestras de vinagre y etanol. Sin embargo, su potencial abarca múltiples áreas, como lo demuestra el estudio de Rahmania et al. (2015) la cuál utilizó esta técnica para la identificación de carne de rata en albóndigas comercializadas en Indonesia. Rahmania et al. En su investigación aplica PLS para cuantificar la carne de rata en muestras y un PCA para clasificar el tipo de carne encontrada. El FTIR permitió diferenciar el tipo de grasa de cada una de las muestras, lo que logró la identificación de la carne adulterada.

El FTIR y sus derivadas también ha sido ampliamente utilizado en la industria alimentaria para la identificación, autenticidad y detección de adulteraciones en productos de consumo. Un ejemplo de ello es el estudio de Sahlan et al. (2019), en el cual se aplicó espectroscopia infrarroja por transformada de Fourier con reflectancia total atenuada (ATR-FTIR) junto con métodos quimiométricos para la identificación y clasificación de la autenticidad de la miel. Los resultados demostraron que ATR-FTIR es una herramienta efectiva y rápida para verificar la autenticidad de la miel y diferenciar su origen, lo que la convierte en una alternativa confiable para el control de calidad en la industria.

De manera similar, (Vinciguerra et al., 2019) utilizó FTIR-ATR para la detección de adulteración en leche con suero de queso, evidenciando la capacidad de esta técnica para identificar fraudes alimentarios y garantizar la calidad de los productos lácteos.

Más allá de su aplicación en el control de calidad, la espectroscopia FTIR también ha demostrado ser una herramienta prometedora en el campo de la salud., un ejemplo claro es la investigación de Smok-Kalwat et al. , quien ocupó el FTIR para detectar cambios químicos en suero y plasma con el fin de indicar la presencia de cáncer de pulmón. Los resultados indicaron que la región espectral más útil para la diferenciación de pacientes con cáncer se encuentra en el rango de $800-1800\text{ cm}^{-1}$.

Smok-Kalwat et al. usó técnicas de aprendizaje automático (machine learning) como análisis de componentes principales (PCA), máquinas de soporte vectorial (SVM), curvas ROC y regresión logística (LR) para evaluar la precisión de FTIR en la detección de cáncer. Este estudio nos muestra el potencial del uso del FTIR para biopsias líquidas, siendo esta una opción eficiente y no invasiva.

Continuando con la demostración de versatilidad del FTIR, nos extendemos a la detección rápida y simultánea de metales pesados en suelos agrícolas como nos demuestra Devianti et al. (2019) en su estudio. En esta investigación, se utilizó espectroscopia de reflectancia infrarroja para identificar la contaminación por zinc (Zn) y plomo (Pb) en suelos afectados por actividades humanas como la construcción, minería, eliminación de aguas residuales, transporte y fertilización.

Devianti et al. en su análisis mediante PCA permitió diferenciar con claridad suelos contaminados y no contaminados, con un 97% de varianza explicada, lo que evidencia una separación efectiva de las muestras según su nivel de contaminación. En el

caso del PLSR, permitió predecir simultáneamente las concentraciones de Zn y Pb con un coeficiente de correlación $(r) = 0.98$ para ambos metales, lo que sugiere una gran precisión en la predicción.

Estos resultados demuestran que el FTIR, puede ser una herramienta eficaz para la monitorización ambiental, proporcionando una solución rápida, sencilla y no invasiva para la detección de contaminantes en suelos agrícolas.

2.2 MARCO TEÓRICO

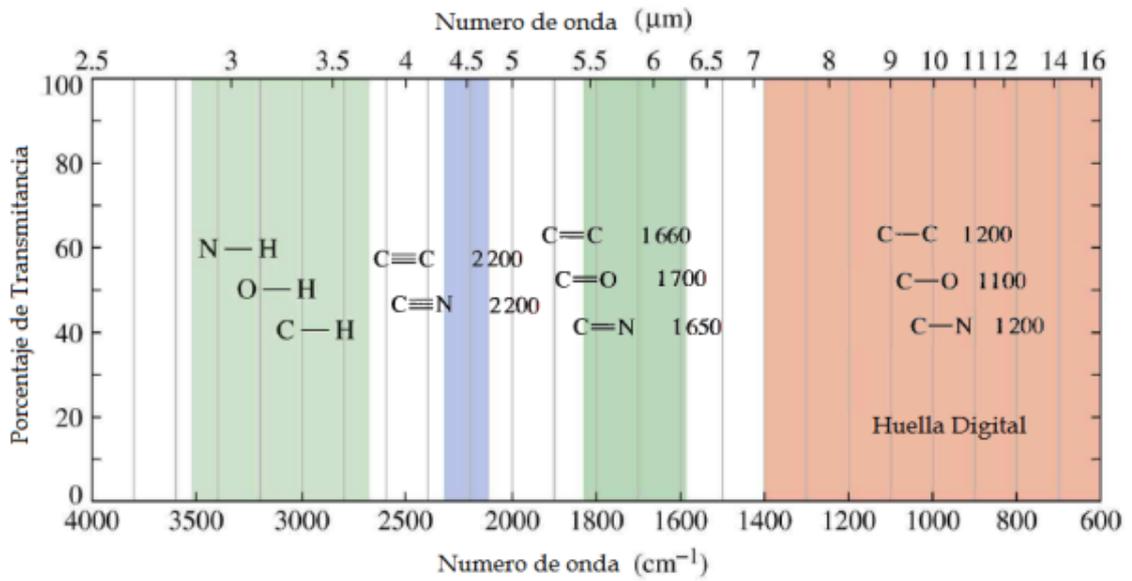
2.2.1 Espectroscopia Infrarroja por Transformada de Fourier (FTIR)

La espectroscopia infrarroja por transformada de Fourier (FTIR) es una metodología empleada en la industria y en los laboratorios académicos para comprender la estructura de las moléculas individuales y la composición de las mezclas moleculares (Athavale et al., 2020). Esta técnica usa la energía del infrarrojo para el análisis de muestras; esta energía pasa a través de la muestra, la cual es absorbida parcialmente, mientras que el restante pasa a través de ella; la señal resultante será un espectro que representa una 'huella digital' molecular de la muestra (FTIR Spectroscopy Academy).

2.2.1.1 Región Diagnóstica y región de las huellas digitales

Los picos que aparecen en la región comprendida entre 1500 y 500 cm^{-1} , se le denomina región de huella digital, la cual son generadas por varias vibraciones de enlace presentes en la estructura de la molécula. (Document, n.d.)

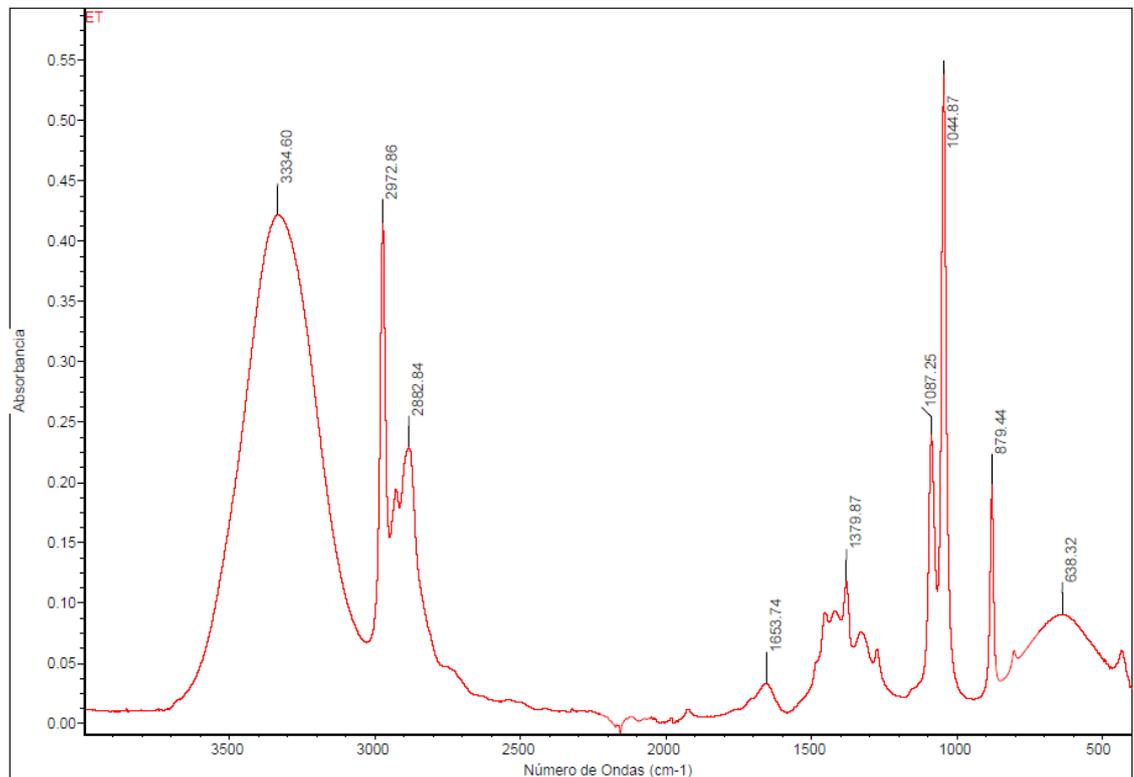
Una vez que tengamos nuestro espectro, analizamos sus grupos funcionales (cada pico generado) y así categorizar la presencia de enlaces que la muestra analizada presenta, es decir, la banda (o pico) que se genera en un espectro se ubica siempre en la misma zona según el grupo funcional que presente. (Document, n.d.)



Fuente: CAPITULO II Espectroscopia Del Infrarrojo 2.1 Región Del Infrarrojo

Basándonos en la gráfica y guiándonos bajo el contexto de este trabajo, las muestras de etanol y ácido acético se encontrarían en rangos aproximados de 1600-1800 cm^{-1} para ácido acético y 1200-1400 para etanol.

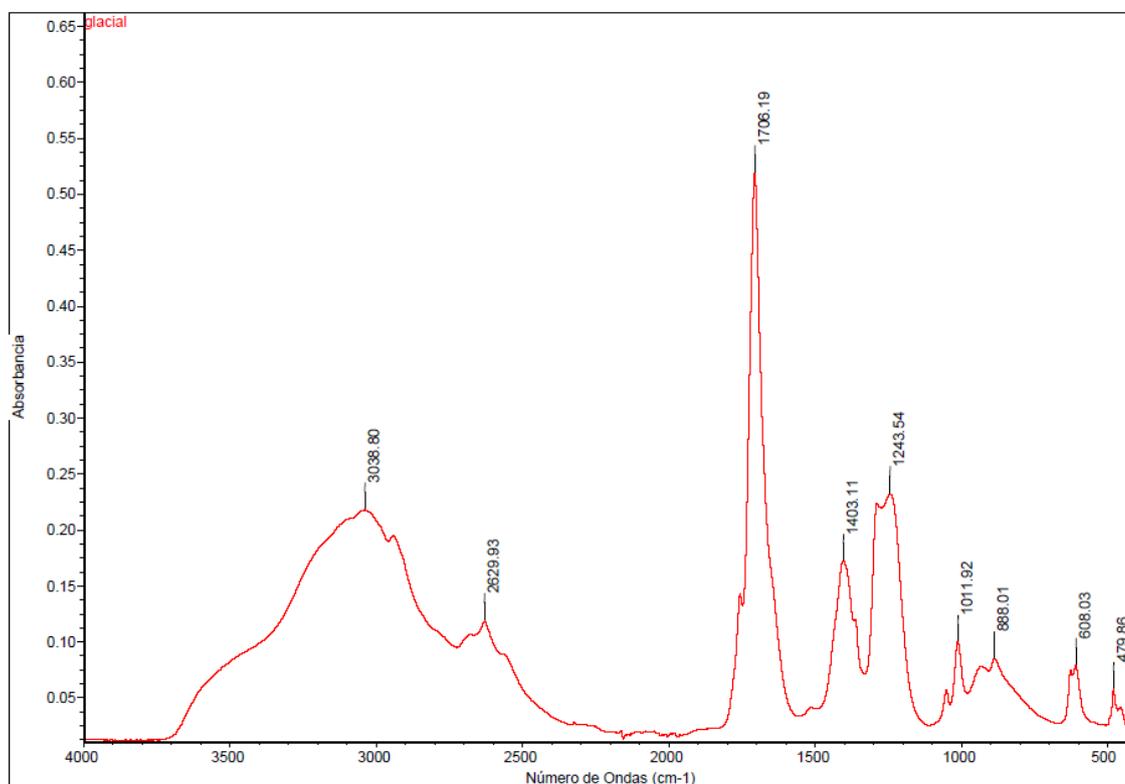
2.2.1.2 Espectro Etanol:



Fuente: Propia autoría

Bajo este ejemplo, podemos observar los picos más relevantes de mi muestra de etanol al 99%, la cual los rangos se encuentran entre 1044 – 1100 cm^{-1} .

2.2.1.3 Espectro Ácido Acético:



Fuente: Propia Autoría

De igual manera, se analizó el ácido acético glacial en su concentración más pura, permitiéndonos la gráfica expuesta, la cual podemos destacar sus picos principales que radican entre el rango de 1243-1706 cm^{-1} .

2.2.1.4 Señal residual del background en el espectro

El análisis del background en un espectro de FTIR consiste en la sustracción automática del espectro de fondo antes de obtener el espectro de la muestra. Este procedimiento permite eliminar las interferencias naturales del ambiente donde se encuentra el haz infrarrojo, asegurando el análisis correcto de las muestras de interés. (CAPITULO II Espectroscopia Del Infrarrojo 2.1 Región Del Infrarrojo, n.d.)

Ahí es donde radica su importancia de realizar un background, antes del análisis de mis muestras.

2.2.2 Diseño Experimental: Diseño de Mezclas

La creación de mezclas es un enfoque estadístico utilizado para analizar y mejorar las características de una combinación de elementos, en la que la respuesta del sistema varía según las proporciones relativas de estos componentes, y la suma total se mantiene constante.(Becerra & Goos, 2021)

Existen diferentes enfoques para estructurar estos experimentos, entre los cuales destacan: el diseño simplex-reticular, el diseño simplex con centroide y el diseño de mezclas con restricciones (Gutiérrez Pulido & de la Vara Salazar, n.d.) En esta investigación, se utilizó el diseño de mezclas con limitaciones, ya que facilita la definición de márgenes concretos en las proporciones de los componentes, garantizando que las combinaciones obtenidas sean representativas y factibles dentro del ámbito del estudio.

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

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

Donde:

x_i : representa la proporción de cada componente en la mezcla

q : es el número total de componentes.

Sin embargo, en algunos casos, los componentes no tienen la capacidad de variar sin restricciones entre 0 y 1, ya que deben ajustarse a ciertas limitaciones del tipo por lo que se sugiere lo siguiente:

$$lc \leq x_i \leq uc$$

Donde x_i representa la proporción para cada analito, lc hace referencia a la mínima restricción de i -ésimo componente, y uc representa la proporción máxima del i -ésimo componente.

2.2.3 Preprocesamiento de Datos Espectrales

2.2.3.1 Corrección de línea base con ALS (Asymmetric Least Squares)

El objetivo de corrección de la línea base es eliminar el fondo para aislar solo los picos importantes.

El principal objetivo del ALS es aplicar diferentes penalizaciones a las desviaciones positivas y negativas al realizar el ajuste, eliminando así picos o dispersiones no deseadas, facilitando mi análisis. (*Two Methods for Baseline Correction of Spectral Data* • NIRPY Research, n.d.)

El método ALS se basa en el suavizado de mínimos cuadrados ponderados asimétricamente. (Peng et al., 2010)

$$z = \mathbf{arg\ min} \sum_i w_i (y_i - z_i)^2 + \lambda \sum_i (\Delta^2 z_i)^2$$

Donde:

y : es el espectro original

w_i : son pesos asimétricos que dependen de la relación entre y_i y z_i

λ : es un parámetro de regularización que controla la suavidad de la línea base

Δ^2 : es el operador de segunda diferencia para imponer suavidad en la línea base.

Para datos espectrales múltiples, se introduce una penalización en las diferencias entre espectros corregidos para mejorar la estabilidad (Peng et al., 2010)

$$z_k = \mathbf{arg\ min} \sum_{u,v} \|(y_u - z_u) - (y_v - z_v)\|^2 + \sum_i w_i (y_i - z_i)^2 + \sum_i \lambda_i \|Dz_i\|^2$$

2.2.3.2 Normalización de datos

Para este estudio se realizó una normalización por área, cuyo objetivo es el de asegurar que todos los espectros tengan la misma área total. Tomando el ejemplo de Hodgkins et al. en su investigación, ocupó este método para equilibrar sus datos químicos de humedad y compensar así las variaciones que la investigación presentó.

Para calcular la normalización, se aplica la siguiente fórmula:

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

Donde:

A_{λ} : es la absorbancia en una longitud de onda específica.

$\sum A_{\lambda}$: es la integral del espectro

$A_{\lambda,norm}$: es la absorbancia normalizada

2.2.3.3 Suavizado y eliminación de ruido

El “ruido” de un espectro de FTIR puede ser reducido mediante un proceso de suavizado(*Document*, n.d.).

El filtro SG es uno de los más utilizados porque preserva mejor la forma de los picos y es ideal para espectroscopía (Savitzky & Golay, 1964). Su 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 : con los coeficientes de convolución obtenidos del ajuste de polinomios de orden k

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

2.2.3.4 Segunda derivada

La segunda derivada de una señal es útil para resaltar características específicas como picos y bordes en espectros. Se usa principalmente en la espectroscopía para la mejorar la resolución y separar picos superpuestos, en los análisis de datos, nos permite detectar los cambios bruscos en la señal. (Savitzky & Golay, 1964)

El filtro de Savitzky-Golay permite calcular la segunda derivada al aplicar una convolución con coeficientes específicos, obtenidos de un ajuste polinómico(Savitzky & Golay, 1964).

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

Donde:

$(c_j)^2$: son los coeficientes de la segunda derivada.

$y_i^{(m)}$: es la segunda derivada del filtro

$y_i + j$: son los valores originales de la señal

2.2.4 Modelamiento Matemático y Quimiometría

2.2.4.1 Análisis de Componentes Principales (PCA)

El PCA resume el contenido informativo de grandes conjuntos de datos en un conjunto más pequeño de variables no correlacionadas conocidas como componentes principales.(IBM, n.d.) Cada componente intenta abarcar la mayor cantidad de información.

En el PCA se calculan dos componentes principales: el primer componente principal (PC1) y el segundo componente principal (PC2).

En el PC1 recoge la mayor varianza en los datos, representando la mejor dirección para proyectar los puntos, en cambio, el PC2 captura la mayor varianza restante ortogonal al PC1, con una correlación de cero entre ambos (IBM, n.d.).

2.2.4.2 Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA):

El Análisis Discriminante de Mínimos Cuadrados Parciales (PLS-DA) es una variante del método Partial Least Squares Regression (PLS) que se utiliza cuando la variable dependiente es categórica. En términos generales, PLS-DA realiza una regresión PLS en una matriz de variables predictoras X para modelar una matriz de variables respuesta Y , donde Y representa categorías binarias o múltiples (Pérez-Enciso & Tenenhaus, 2003a)

A diferencia del análisis discriminante clásico, PLS-DA es particularmente útil cuando el número de predictores es mucho mayor que el número de observaciones y cuando existe alta multicolinealidad, condiciones comunes en el análisis de datos espectrales y en estudios de quimiometría (Pérez-Enciso & Tenenhaus, 2003a). Esta técnica ha sido ampliamente aplicada en el análisis de microarreglos para la clasificación

de tejidos en estudios biomédicos, así como en la discriminación de grupos en estudios de composición química de alimentos y otros sistemas complejos (Nguyen & Rocke, 2002)

El PLS-DA se ha utilizado en múltiples campos, incluyendo:

- **Clasificación de muestras biológicas:** En estudios genéticos, PLS-DA ha permitido discriminar entre tejidos sanos y tumorales basándose en datos de microarreglos (Nguyen & Rocke, 2002).
- **Análisis de datos espectrales:** En espectroscopía FTIR y Raman, PLS-DA es útil para la clasificación de compuestos con estructuras químicas similares, ayudando a identificar contaminantes en mezclas complejas (Pérez-Enciso & Tenenhaus, 2003a).

2.2.4.3 Sparse PLS-DA

sPLS-DA es una técnica de clasificación basada en PLS que es especialmente útil en datos con muchas variables y en problemas multiclase. La versión sPLS-DA permite seleccionar las características más relevantes de manera automática para mejorar la clasificación y evitar sobreajuste. (Fisher, 1936)

2.2.4.4 Curva ROC (Receiver Operating Characteristic):

La Curva de Característica Operativa del Receptor (ROC, por sus siglas en inglés) es una herramienta gráfica utilizada para evaluar el rendimiento de un modelo de clasificación binaria. Su principal ventaja es que permite visualizar el balance entre la sensibilidad (true positive rate, TPR) y la especificidad (true negative rate, TNR) de un clasificador a diferentes umbrales de decisión. (Luo et al., 2012)

En la práctica, la curva ROC es ampliamente utilizada en:

- Diagnóstico biomédico
- Quimiometría
- Espectroscopia y clasificación química
- Evaluación de modelos de aprendizaje automático

La curva ROC se construye a partir de dos métricas clave en la clasificación binaria:

Sensibilidad (True Positive Rate, TPR): mide la capacidad del modelo para identificar correctamente las muestras positivas.

$$\text{Sensibilidad} = \frac{TP}{TP + FN}$$

Donde:

TP: son los positivos correctamente clasificados.

FN: son los positivos incorrectamente clasificados como negativos.

Especificidad (True Negative Rate, TNR): mide la capacidad del modelo para identificar correctamente las muestras negativas.

$$\text{Especificidad} = \frac{TN}{TN + FP}$$

Donde:

TN: son los negativos correctamente clasificados.

FP: son los negativos incorrectamente clasificados como positivos.

El Área Bajo la Curva ROC es un valor numérico que cuantifica el rendimiento del clasificador:

AUC = 1.0: Clasificador perfecto.

AUC > 0.9: Clasificador excelente.

AUC entre 0.7 y 0.9: Clasificador aceptable.

AUC ≈ 0.5: Clasificación aleatoria.

En espectroscopía, una AUC alta indica que la selección de variables espectrales es eficiente para distinguir clases (Luo et al., 2012)

2.2.4.5 Regresión por Mínimos Cuadrados Parciales (PLS) en análisis cuantitativo.

El método de Mínimos Cuadrados Parciales (PLS, Partial Least Squares) es una técnica de modelado multivariado que se utiliza cuando hay un gran número de variables predictoras (X) altamente correlacionadas y un conjunto de variables respuesta (Y). PLS es ampliamente utilizado en análisis quimiométrico, espectroscópico y en la bioinformática para la selección de variables y reducción de dimensionalidad (Lê Cao et al., 2011).

A diferencia del Análisis de Componentes Principales (PCA), que solo maximiza la varianza de X , el PLS busca encontrar componentes latentes que maximicen la covarianza entre X y Y , lo que lo hace útil en problemas de regresión y clasificación (Lê Cao et al., 2011).

El modelo PLS se expresa como:

$$Y = X\beta + E$$

Donde: β es la matriz de coeficientes de regresión obtenida a partir de las componentes latentes, y E es la matriz de residuos (Lê Cao et al., 2011).

2.2.4.6 Sparse PLS

El método Sparse PLS (sPLS) es una extensión de PLS que incorpora un mecanismo de selección de variables mediante penalización, lo que permite identificar solo las variables más relevantes para la predicción. Fue desarrollado para abordar problemas de alta dimensionalidad, como los que se encuentran en estudios de genómica, quimiometría y espectroscopía (Lê Cao et al., 2011).

2.2.4.7 RMSE

El Error Cuadrático Medio de la Raíz o Root Mean Square Error (RMSE) es una métrica ampliamente utilizada para evaluar la precisión de modelos de calibración en análisis espectroscópico, especialmente en técnicas como Partial Least Squares (PLS). Este error cuantifica la diferencia promedio entre los valores predichos y los valores reales, proporcionando una medida de la capacidad del modelo para realizar predicciones precisas.

El RMSE se calcula utilizando la siguiente ecuación:

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

Donde y_i representa los valores reales, \hat{y}_i los valores predichos por el modelo, y n el número total de muestras.

Un valor de RMSE menor indica una mejor precisión del modelo, es decir, una menor desviación entre los valores predichos y los valores reales. (Christin Brettschneider et al., 2022)

2.3 MARCO CONCEPTUAL

- **Análisis de Componentes Principales (PCA):** Descripción de esta técnica para la reducción de dimensionalidad y simplificación de datos.
- **Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA):** técnica estadística que combina el PLS para clasificar datos altamente correlacionados, optimizando la separación de clases en problemas de alta dimensionalidad.
- **Contribución de Variables:** importancia de cada predictor en la clasificación, evaluada mediante pesos, cargas, coeficientes de regresión y score, para identificar las más relevantes.
- **Espectro FTIR:** Señales obtenidas corresponden a las vibraciones moleculares específicas.
- **Mínimos Cuadrados Parciales (PLS):** Explicación de esta técnica estadística utilizada para crear un modelo predictivo entre los datos espectrales y las concentraciones de analitos.
- **Quimiometría:** Definición de la disciplina que utiliza técnicas matemáticas y estadísticas para interpretar datos químicos.

CAPÍTULO 3

MATERIALES Y MÉTODOS

3.1 NIVEL DE INVESTIGACIÓN

La presente investigación es de tipo explicativa, debido a que se analiza la relación de causa y efecto entre la contaminación ácido acético y los diferentes picos que se obtienen a través de la espectroscopia infrarroja por transformada de Fourier (FTIR).

3.2 DISEÑO DE INVESTIGACIÓN

Este trabajo sigue un diseño de investigación experimental, ya que las muestras de ácido acético son sometidas a diversas concentraciones controladas de etanol. De este modo, es posible observar la relación de causa y efecto mencionada anteriormente.

3.3 DISEÑO DE EXPERIMENTO

Para generar la base de datos de espectros necesaria para entrenar los modelos predictivos, se lleva a cabo un diseño de mezclas con restricciones. Este enfoque permite la creación estándares representativos con diferentes concentraciones de los analitos de interés, en este caso el etanol y el ácido acético, ajustándose a contextos reales en los que la presencia de estos compuestos está limitada por razones técnicas y de control. Dicho diseño garantiza que las concentraciones de los analitos reflejan condiciones realistas del proceso de producción de vinagre.

El diseño de mezclas se expresa de la siguiente forma:

$$0 \leq lc \leq xi \leq uc \leq 1$$

Para los fines de esta investigación se establecen los rangos:

$$0 \leq 5 \leq \acute{A}c. \acute{A}c\acute{e}t\acute{i}c\acute{o} \leq 6 \leq 1$$

$$0 \leq 0.5 \leq Etanol \leq 1 \leq 1$$

Este rango asegura que las concentraciones de etanol y ácido acético abarquen los valores observados típicamente en el vinagre comercial y se encuentren dentro de los límites normativos. Las concentraciones de los analitos representados previamente serán

expresadas en porcentaje para su análisis. Para generar las mezclas óptimas bajo las restricciones establecidas, se utiliza el software R junto con los paquetes especializados como son: *rms* y *dplyr*.

A continuación, se presenta una tabla de las mezclas con su respectivo código, la imagen muestra una vista parcial de la tabla (se podrá encontrar la tabla completa en el compartimiento de ANEXO 1), por lo que es importante mencionar, que se trabajará con 25 muestras contaminadas, y 10 muestras con distintas concentraciones de los analitos por separado, lo que da un total de 35 muestras a analizar mediante FTIR, con 3 réplicas de cada uno.

AcAcético	Etanol	Agua	Código
5	0,5	0,945	M1
5,25	0,5	0,9425	M2
5,5	0,5	0,94	M3
5,75	0,5	0,9375	M4
6	0,5	0,935	M5
5	0,625	0,94375	M6
5,25	0,625	0,94125	M7
5,5	0,625	0,93875	M8
5,75	0,625	0,93625	M9

Tabla 1: Fragmento de mezclas (1-9) del diseño experimental

Fuente: Propia autoría

Para complementar la información, se generaron dos tipos de gráficos para mejor interpretación de datos, uno de dispersión y otro interactivo de superficie en 3D. Estos fueron elaborados con los paquetes *ggplot2* y *plotly*, respectivamente, para visualizar de manera más efectiva el diseño experimental.

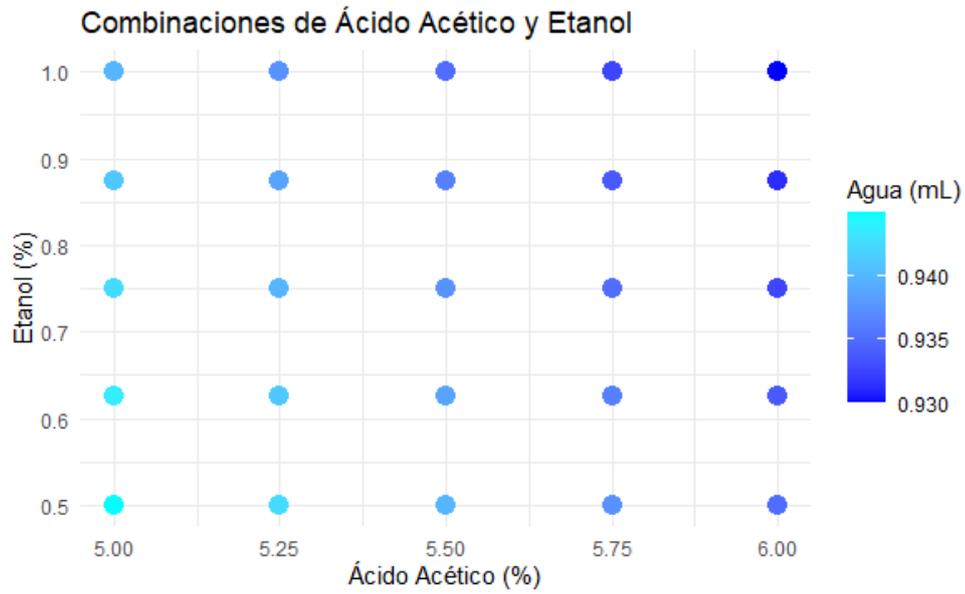


Figura 2: Gráfico de dispersión de las mezclas generadas en R

Fuente: Propia autoría

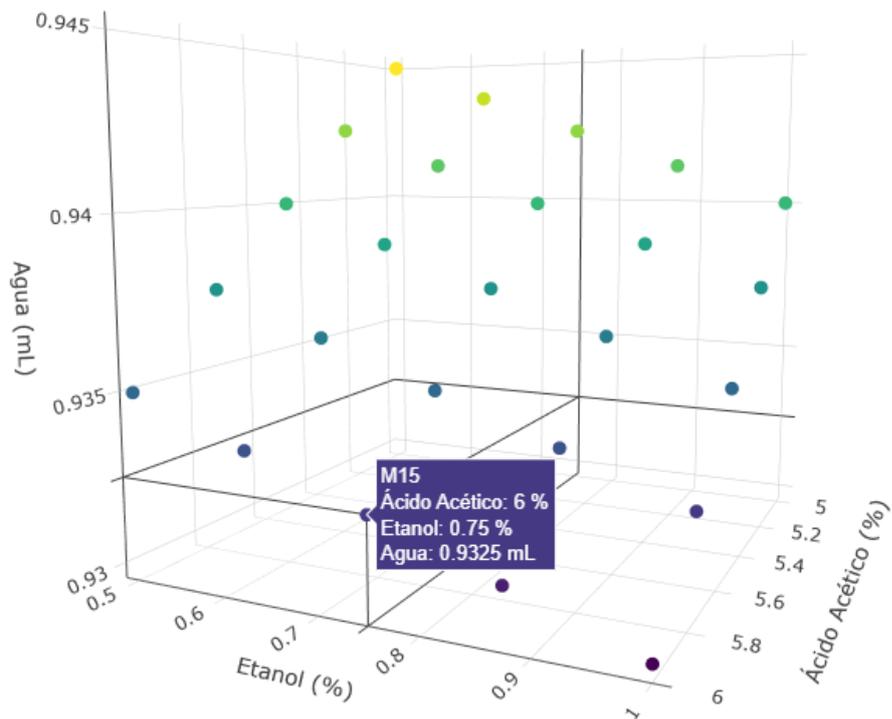


Figura 1: Gráfico interactivo de Superficie en 3D del diseño experimental, incluido una etiqueta detallando los datos por muestra

Fuente: Propia autoría

3.4 VARIABLES

3.4.1 VARIABLES INDEPENDIENTES

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

VARIABLE	DIMENSIÓN	INDICADOR
Longitudes de Onda	Número de Longitudes de Onda Seleccionadas	Unidad de Medida
		Absorbancia
		Rango espectral (400 a 4000 cm^{-1})

Fuente: Propia autoría

3.4.2 VARIABLES DEPENDIENTES

Variable bidimensional (Ác. Acético - Etanol).

VARIABLE	DIMENSIÓN	INDICADOR
Concentraciones	N*2	Unidad de concentración (% v/v)
	N=número total de muestras	

Fuente: Propia autoría

3.4.3 VARIABLES INTERVINIENTES

- Condiciones de preparación de la muestra.
- Volumen de la muestra
- Analista
- Tiempo de almacenamiento del vinagre antes de la medición.

3.4.4 VARIABLES EXTRAÑAS

Entre las variables extrañas que puedan afectar el proyecto están:

- Humedad del ambiente donde se realizan las mediciones.
- Temperatura durante el análisis.
- Pureza de los reactivos utilizados para la elaboración de las muestras.
- Fallas en el equipo FTIR (cualquier mal funcionamiento o desajuste).

3.5 TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Las técnicas que se emplearon son análisis documental y análisis de contenido, cuya recolección de información se realizó a través de distintas fuentes o bases de datos como SRPINGER, SCOPUS, SCHOLAR GOOGLE y repositorios universitarios, entre otros.

Por otro lado, en el diseño de campo, se utilizó la técnica de observación estructurada, donde los espectros obtenidos mediante FTIR serán registrados en archivos con formato .CSV para su posterior análisis en R.

3.6 TÉCNICAS DE PROCESAMIENTO Y ANÁLISIS DE DATOS

Para el preprocesamiento de datos, se trabajó directamente en R para asegurar un mejor control, una vez obtenido los datos se emplearon técnicas como corrección de línea base (ALS), suavización (SG), normalización y segunda derivada.

Posteriormente, se realizó un preprocesamiento a los datos para luego aplicarlo los métodos de regresión multidimensional y reducción de dimensionalidad, incluyendo: PCA (Análisis de Componentes Principales), PLS-DA (Análisis Discriminante por

Mínimos Cuadrados Parciales) y sPLS (Regresión por Mínimos Cuadrados Parciales Escasa).

El desarrollo del procesamiento y análisis de datos se llevó a cabo en el entorno de programación R, en conjunto a archivos en formato .CSV donde se almacenó la información generada.

3.7 PROTOCOLO A IMPLEMENTAR

El siguiente apartado describe el procedimiento a seguir en esta investigación.

3.7.1 FASE PREEXPERIMENTAL

Si bien la fase preexperimental puede parecer sencilla, cada acción realizada es fundamental para garantizar la obtención de resultados precisos y confiables.

3.7.1.1 DISEÑO Y PREPARACIÓN DE MEZCLAS

Como se mencionó anteriormente, se estableció un diseño experimental con restricciones basado en las normativas de CODEX, según las cuales la concentración de ácido acético debe mantenerse en el rango de 5-6% y la de etanol entre 0.5-1% si trabajamos con muestras de vinagre. La tabla generada en R muestra estas concentraciones expresadas en porcentaje (% v/v).

Para el desarrollo de las mezclas, se trabajó con los analitos en su forma más pura, asegurando un manejo adecuado para evitar contaminaciones. Se realizaron diluciones controladas para luego preparar las mezclas correspondientes. En total se generaron 35 muestras, distribuidas de la siguiente manera: 10 muestras estándar con concentraciones específicas del ácido acético y etanol, las cuales se las ocupa como referencia posteriormente; y 25 muestras contaminadas como se muestra en la **¡Error! No se encuentra el origen de la referencia..**

3.7.1.2 REGISTRO Y ORGANIZACIÓN DE DATOS EN R

Las mezclas obtenidas se almacenaron en una base de datos en formato .CSV para su análisis. La carga de estos datos en R se llevó a cabo mediante los paquetes *ir* y *dplyr*, facilitando el proceso de lectura de datos.

3.7.2 FASE EXPERIMENTAL

En esta fase se abarcará la ejecución del experimento y la adquisición de los espectros analizados por el FTIR.

3.7.2.1 Preparación y medición de muestras en FTIR

Las muestras fueron almacenadas en tubos Eppendorf de 1.5 ml. A cada análisis se le asignó un código único, indicando el número de réplica correspondiente. Antes de cada medición, las muestras fueron homogeneizadas por 10 segundos mediante un vortex para garantizar su uniformidad antes de ser analizadas en el equipo FTIR. Cada muestra fue analizada en tres réplicas, depositando 55 μL en la zona de medición del espectrofotómetro infrarrojo.

3.7.2.2 Adquisición y almacenamiento de espectros

Durante la adquisición de espectros, se llevó a cabo un análisis de background cada 20 minutos para mantener la precisión del equipo. Es importante destacar que los datos obtenidos se trabajaron en su forma cruda, sin aplicar modificaciones en el software OMNIC.

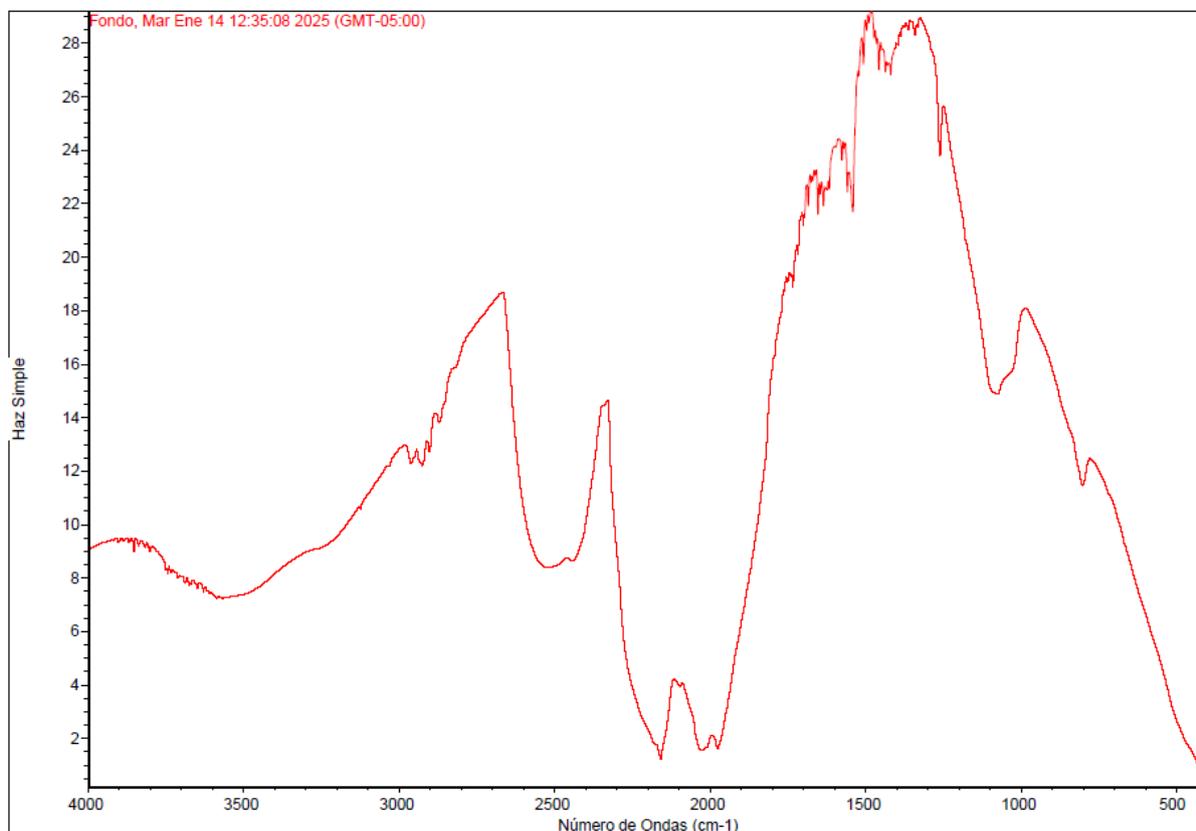


Figura 3: Espectro del fondo del laboratorio de cromatografía

Fuente: Propia autoría

Finalmente, tras cada análisis, los espectros fueron almacenados en formato .CSV para su posterior procesamiento y análisis en R.

3.7.2.3 Preprocesado de datos

Antes de la construcción del modelo, los datos fueron sometidos a un preprocesamiento que incluyó corrección de línea base, normalización, suavizado y segunda derivada. En la siguiente figura, podremos observar espectros de cada color, la cual indicará cada una de las muestras contaminadas controladas con sus respectivas réplicas.

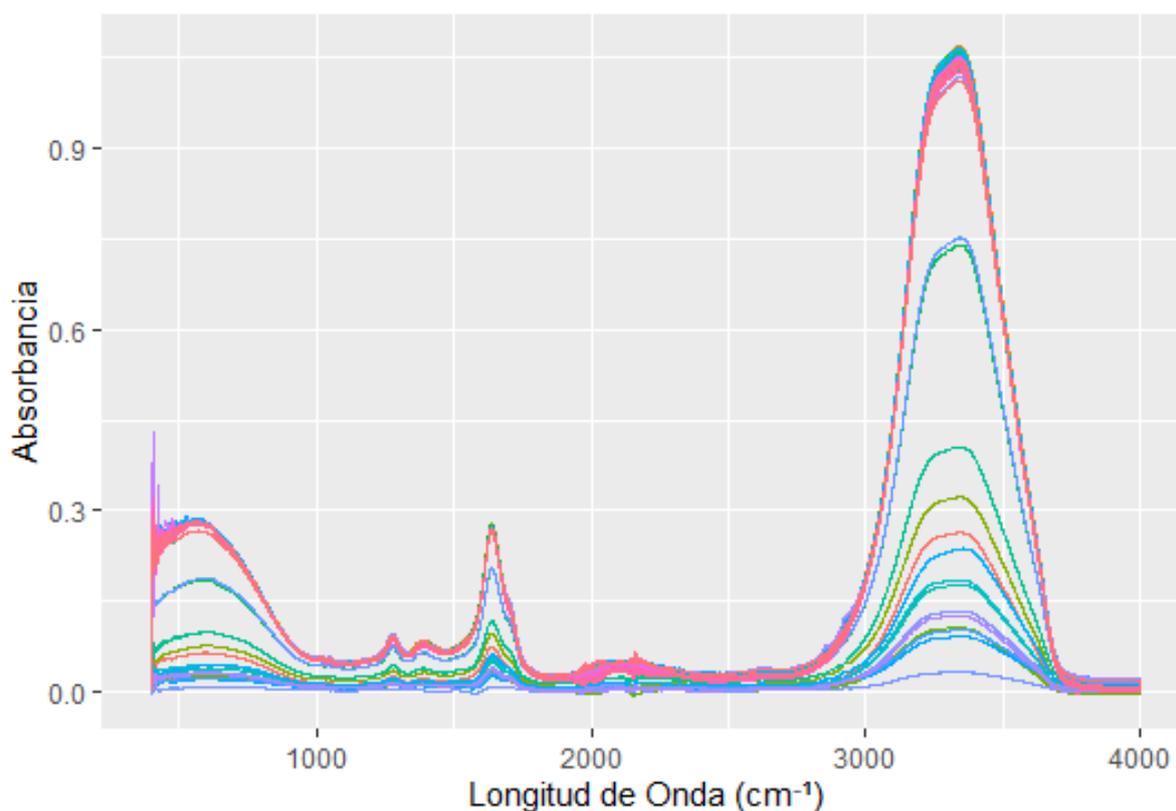


Figura 4: Espectros sin procesar de muestras contaminadas controladas

Fuente: Propia autoría

3.7.2.3.1 Corrección de línea Base

Para la corrección de línea base se ocupó el método ALS, la cual aplica una penalización asimétrica que minimiza la contribución de la línea base sin afectar los picos de interés (Peng et al., 2010); este método no solo ajusta la línea base, sino que también ayuda a eliminar interferencias no deseadas (Peng et al., 2010), como el pico de enlaces OH (agua), permitiendo que el modelo tenga un mejor filtrado y conserve únicamente las señales más relevantes de cada analito. Para aplicar este método se utilizó el paquete “*baseline*” empleando la función

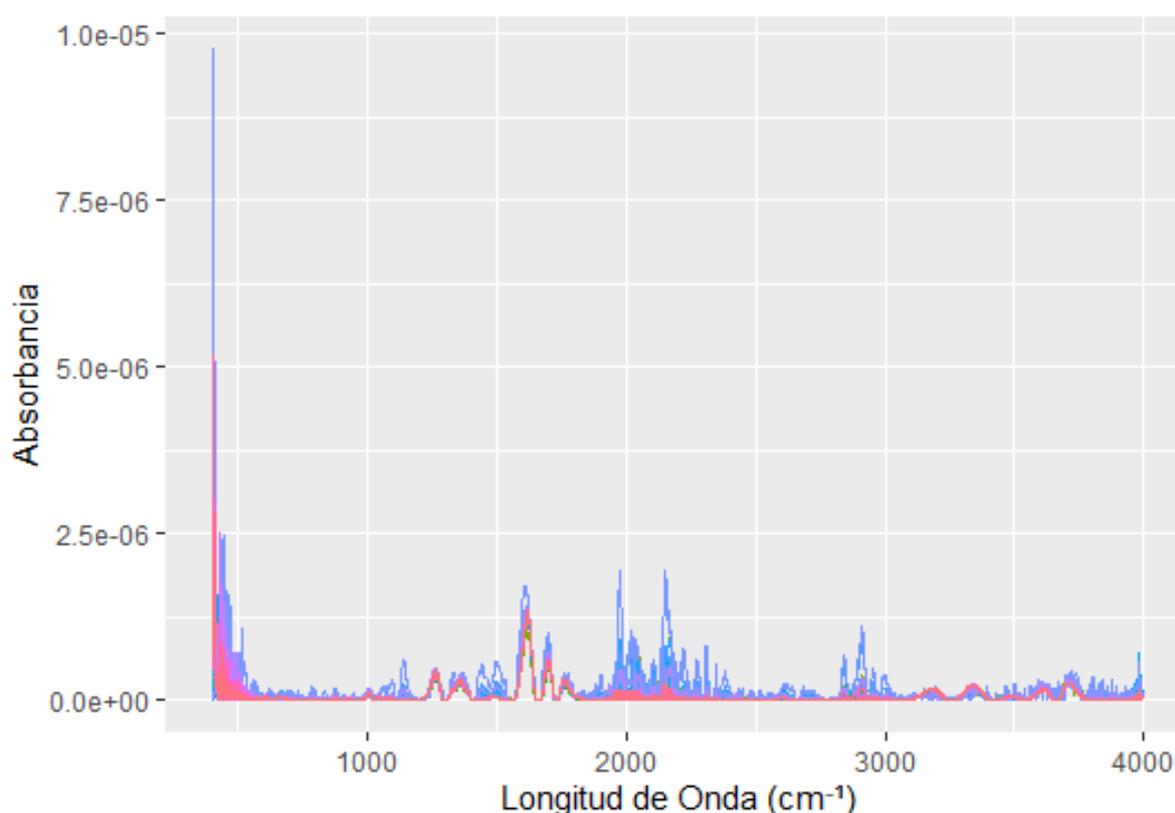


Figura 5: Espectro de muestras contaminadas controladas después de la corrección de línea base (ALS)

Fuente: Propia autoría

3.7.2.3.2 Normalizado

Después de la corrección base, los espectros fueron sometidos a normalización de los valores de intensidad dividiendo cada valor de intensidad por la suma de todos los valores de intensidad. (Henning Teickner, 2018). Para su aplicación en R, se usó una función que se muestra en **¡Error! No se encuentra el origen de la referencia.** en el script bajo el nombre de TESIS FINAL, en la línea (902-906)

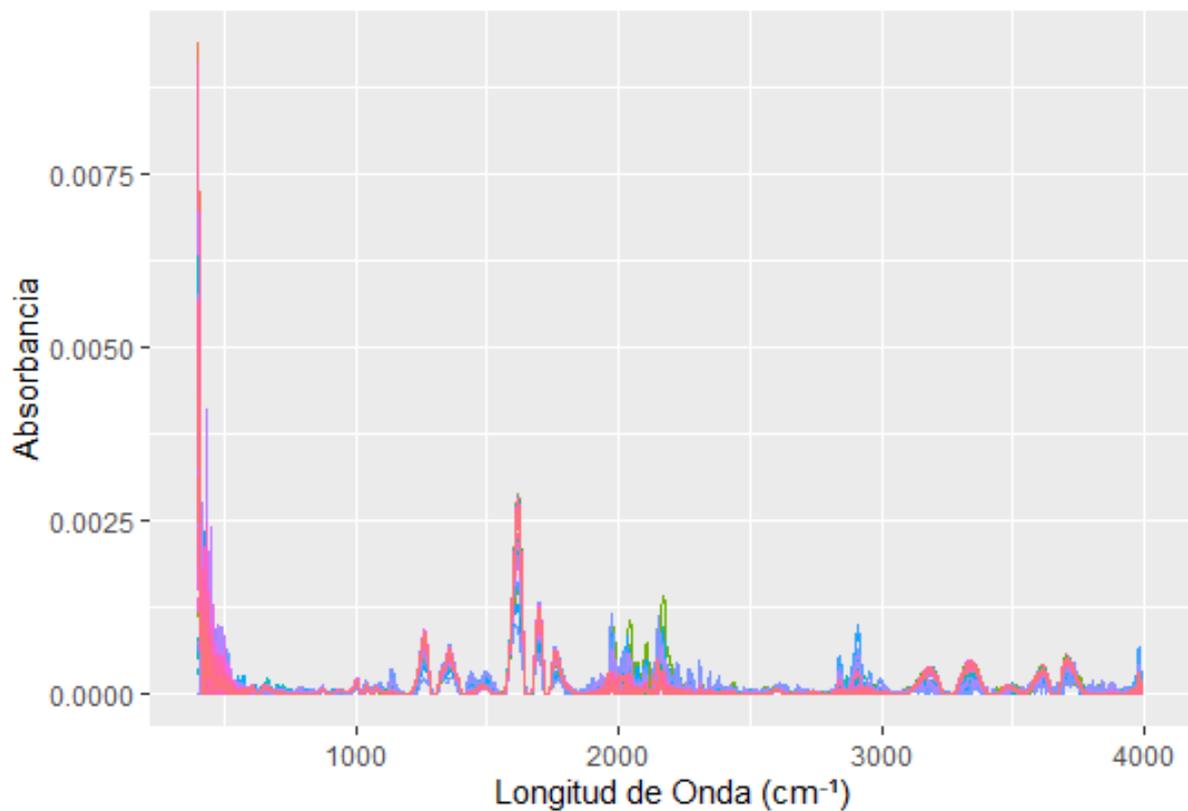


Figura 6: Espectro del normalizado por área de las muestras contaminadas controladas

Fuente: Propia autoría

3.7.2.3.3 Suavizado

Para el suavizado del espectro, se siguió la metodología propuesta por Henning Teickner (2018), utilizando el algoritmo de Savitzky-Golay; la cual nos ayudó a reducir el ruido del espectro sin distorsionar la forma de los picos espectrales. Para su implementación en R, se utilizó el siguiente comando: línea (163-179; 902-906)

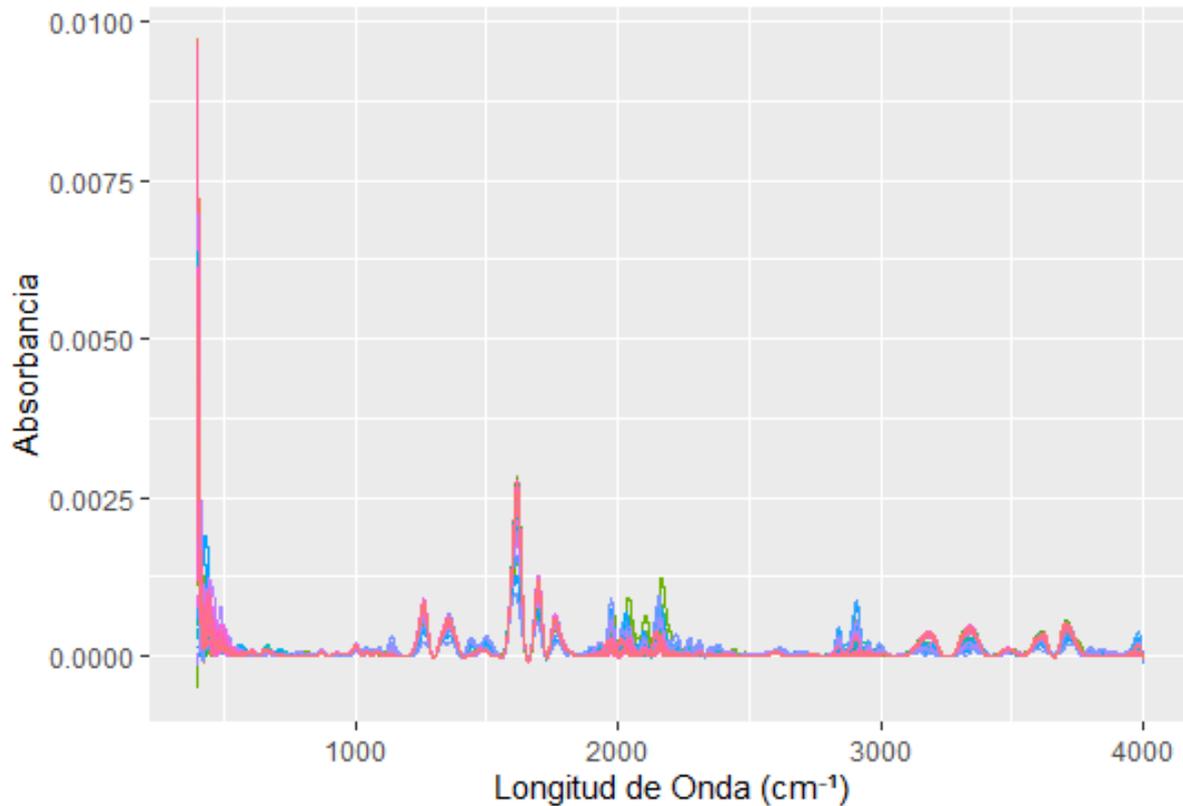


Figura 7: Suavizado del espectro de las muestras contaminadas controladas

Fuente: Propia autoría

3.7.2.3.4 Segunda Derivada

Para calcular la segunda derivada se siguieron las indicaciones de Henning Teickner (2018) , la cual propone ocupar el filtro de Savitzky-Golay. Sin embargo, a diferencia del enfoque del autor en conjunto del paquete “*ir*”, en este estudio se optó por ocupar el siguiente comando bajo el paquete de “*signal*”, la cual aplicará el filtro fila por fila:

Este comando nos brinda libertad para modificar las variables de p , n y m según sea el caso.

- p : representa el orden del filtro, es decir, el grado del polinomio utilizado en el ajuste.
- n : indica la cantidad de valores de número de onda utilizados para construir el polinomio.
- m : corresponde a la derivada m -ésima que se desea calcular.

Bajo ese contexto, ocupamos la siguiente combinación ($p = 2$, $n = 7$, $m = 2$) para calcular la segunda derivada. Línea (923-928)

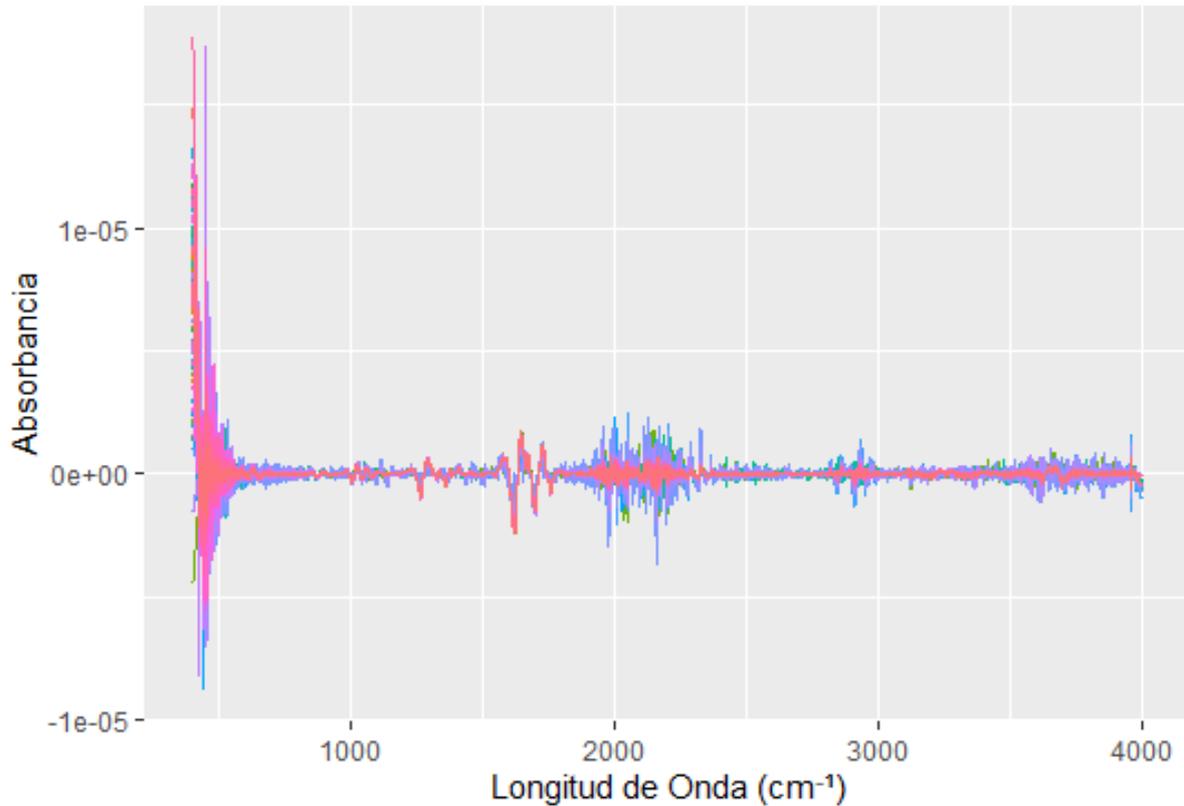


Figura 8: Aplicación de la segunda derivada al espectro de muestras contaminadas controladas

Fuente: Propia autoría

3.7.3 MÉTODOS ESTADÍSTICOS

3.7.3.1 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

El Análisis de Componentes Principales (PCA) es una técnica estadística de reducción de dimensionalidad de los datos espectrales cuyo propósito principal es evaluar la variabilidad de los espectros y determinar si existe una diferenciación clara entre los analitos de interés (ácido acético y etanol). (IBM, n.d.)

Para la implementación del PCA se trabajó solo con los estándares (muestras individuales de ácido acético y etanol por separado), se busca asegurar que se trabajó con analitos sin presencia de contaminación.

Todos los datos utilizados en PCA fueron previamente preprocesados. A continuación, se presentan los códigos empleados para el cálculo del PCA, en la línea (514-533).

A partir del PCA, se generaron gráficos de distribución y sedimentación de los estándares con el propósito de mejorar el análisis y aplicar un filtro adecuado a los datos.

3.7.3.1.1 Contribución de Variables

Para la generación del gráfico de contribución de variables, se utilizó el paquete *factoextra* en R, cuya aplicación lo encontrará en la línea (557-562).

Este gráfico permite identificar qué longitudes de onda aportan más información a la variabilidad de los datos y cuáles son menos relevantes de la PCA.

3.7.3.1.2 Sedimentación de Componentes

El gráfico de Sedimentación o Scree Plot, hace referencia a la cantidad de varianza explicada por cada componente principal en el análisis PCA. El gráfico generado ayuda a determinar cuántos componentes principales se deben considerar en el análisis sin perder información relevante. Lizarazo et al. (2022). El comando ocupado se puede observar en la línea (581-583).

3.7.3.2 ANÁLISIS DISCRIMINANTE POR MÍNIMOS CUADRADOS PARCIALES (PLS-DA)

El Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA) es una técnica de clasificación supervisada, su objetivo es encontrar una relación entre las variables espectrales y las clases predefinidas, permitiendo maximizar la separación entre los grupos.(Pérez-Enciso & Tenenhaus, 2003b)

Para este proyecto, se utilizaron los estándares (analitos por separado) para evaluar si las características espectrales de cada analito presentan diferencias significativas que faciliten su clasificación.

3.7.3.2.1 División en Conjuntos de Entrenamiento (Training) y Prueba (Test)

Para la implementación del PLS-DA, los datos seleccionados fueron divididos aleatoriamente en dos conjuntos: entrenamiento (70%) y prueba (30%), asegurando que

la distribución de los datos fuera representativa y balanceada. Así controlamos el modelo y evitamos un sobreajustamiento del mismo.

3.7.3.2.2 Cálculo del modelo de PLS-DA

Para la generación del modelo sPLS-DA se computó con el paquete de “*mixOmics*”. Este paquete sirve para calcular el sparse PLS-DA, la cual realiza la selección y clasificación de variables en un procedimiento de un solo paso.-(Fisher, 1936)

Posterior al cálculo del modelo, se lo entrenó configurando su número óptimo de componentes (*ncomp*) y las variables seleccionadas (*keepX*), mediante el grid search con Mfold, este modelo entrenado permite identificar patrones en los datos espectrales que diferencian los estándares, facilitando su clasificación.

3.7.3.2.3 Predicción del Modelo PLS-DA

Una vez entrenado el modelo, se realizaron predicciones sobre el conjunto de prueba utilizando la distancia de *Mahalanobis* para asignar cada muestra a su clase correspondiente. El uso de Mahalanobis nos permite determinar la similitud entre dos variables aleatorias multidimensionales, teniendo en cuenta la correlación entre ellas.(Lorenzo, n.d.)

3.7.3.2.4 Curva ROC

A partir del modelo entrenado del PLS-DA, se evaluó su desempeño mediante el cálculo de la curva ROC (Receiver Operating Characteristic). Para este análisis, se consideró un solo componente principal (*ncomp = 1*), porque es la componente que separaba las clases.

3.7.3.3 REGRESIÓN POR MÍNIMOS CUADRADOS PARCIALES (sPLS)

Para el modelamiento, se realizó un proceso similar de en la división de los datos en entrenamiento y prueba que se realizó en el PLS-DA, la diferencia entre el PLS-DA y el sPLS es que los datos fueron divididos en una proporción distinta, mientras en el sPLS-DA se ocupó una relación 70:30, en sPLS se utilizó una relación 60:40 para optimizar el ajuste del modelo.

Además, otra diferencia clave en el enfoque del sPLS es que, en lugar de usar los estándares, el modelo se entrenó con las muestras contaminadas controladas.

Para mejorar la precisión del modelo, se ajustaron sus parámetros mediante la modificación de los rangos de análisis. Con este propósito, se desarrolló una función bucle

(líneas 679-699) que evalúa diferentes combinaciones de parámetros a través de secuencias repetidas, permitiendo determinar la configuración óptima para el modelo final.

Para el cálculo del modelo, se ocupó la función que se ocupa en la línea (186-231) para calcular el sPLS, ya que la variante dispersa permite la selección de las características más predictivas o discriminativas en los datos.(Fisher, 1936). Además, se realizaron dos gráficas, una de contribución y otra de bloque de componentes principales que serán analizadas en la siguiente sección.

3.7.3.3.1 RMSE: Root Mean Square Error (Error Cuadrático Medio)

Después de obtener el modelo, se procedió a evaluar su desempeño mediante el cálculo del Error Cuadrático Medio (RMSE) para sus dos componentes principales. Para ello, se desarrolló una función que implementa la ecuación del RMSE, proporcionando los valores de error en formato de lista, donde cada elemento representa el RMSE de un analito en una componente específica.

3.7.3.3.2 Predicción de muestras de vinagre de manzana

Para culminar, se evaluó la capacidad predictiva del modelo generado en la estimación de las concentraciones de ácido acético y etanol en dos muestras de vinagre de manzana: una de elaboración casera y otra adquirida en supermercado.

Primero, se obtuvo los espectros de las dos muestras de vinagre, cada uno con una etiqueta; a la muestra casera se la denominó “madre” y a la adquirida en supermercado, se la denominó “comercial”.

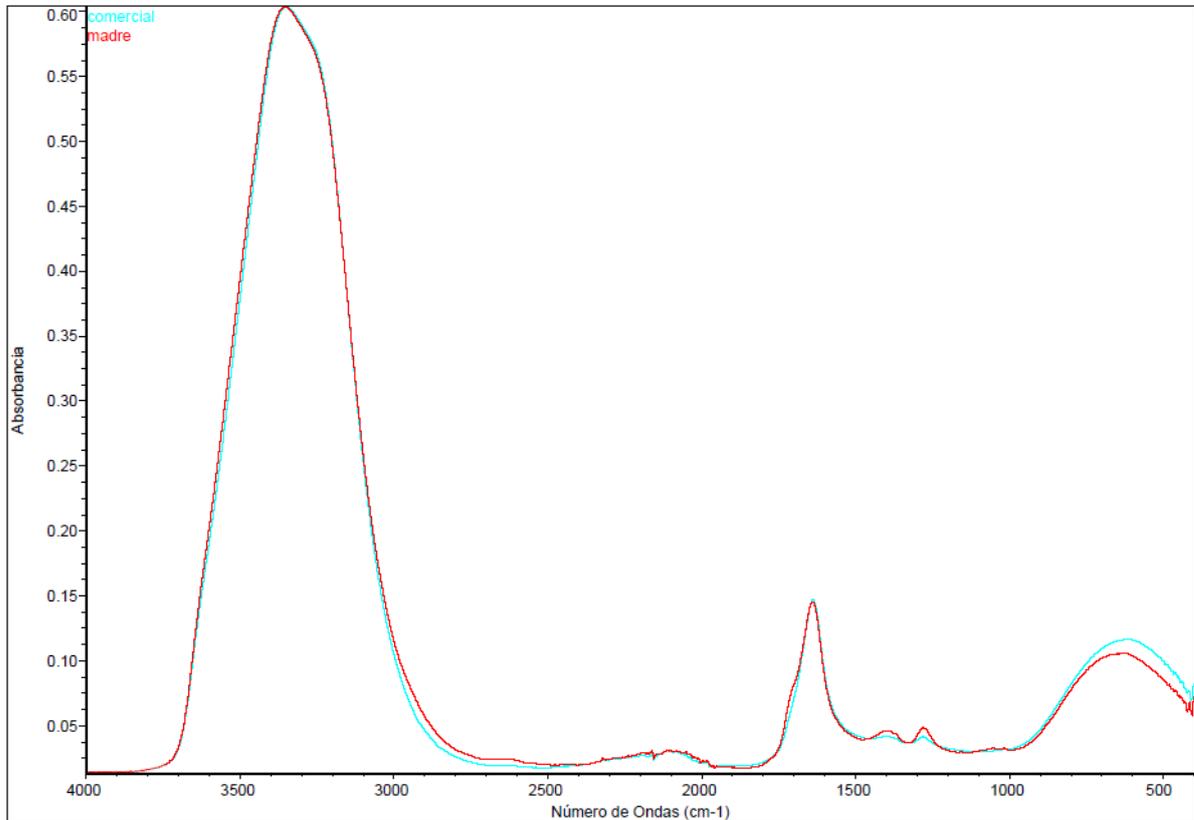


Figura 9: Espectro de las muestras de vinagre de manzana previo a su preprocesamiento

Fuente: Propia autoría

Una vez obtenido los espectros, se importó al R para aplicar el mismo preprocesamiento utilizado en el modelamiento, la cual incluye:

- Corrección de línea base mediante ALS,
- Normalización,
- Suavizado,
- Aplicación de la segunda derivada mediante el filtro Savitzky-Golay.

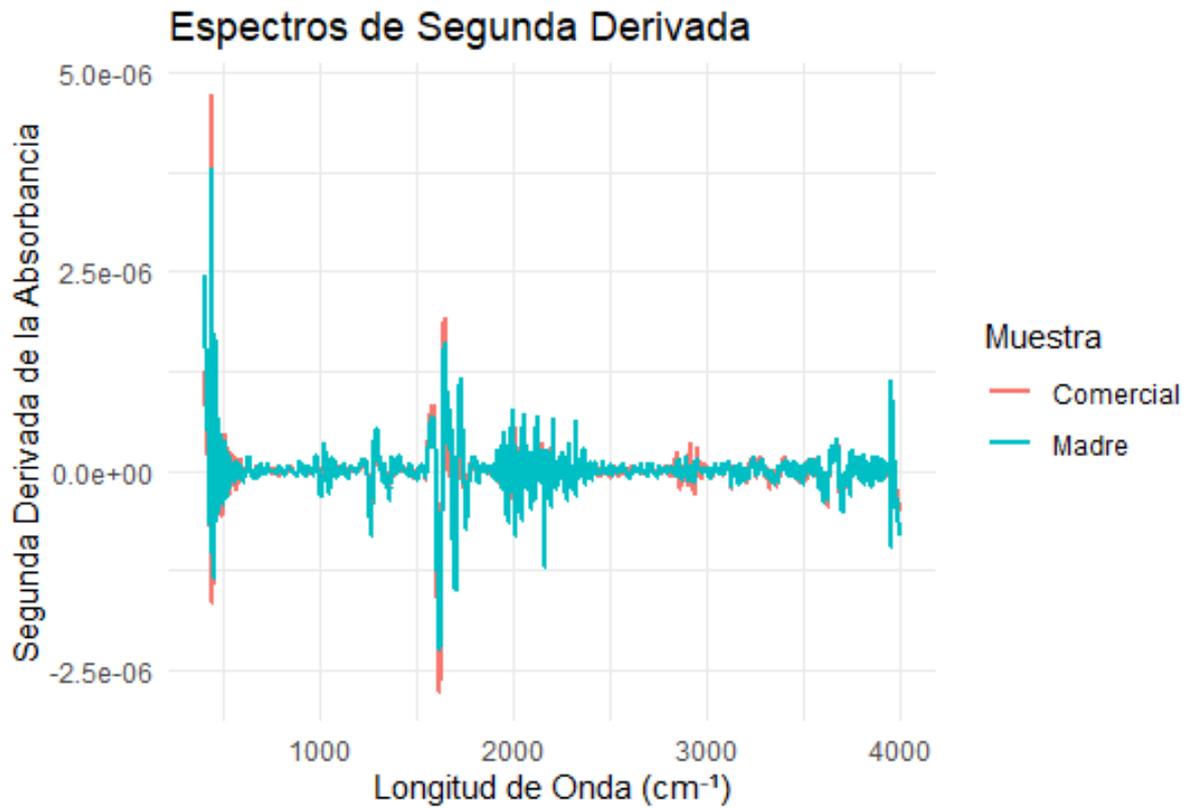


Figura 10: Espectro de las muestras de vinagre de manzana posterior a su preprocesamiento

Fuente: Propia autoría

Finalmente, para la predicción, se utilizó el argumento *newdata* de la función, que corresponde a la matriz espectral preprocesada obtenida a partir de las muestras de vinagre de manzana.

CAPÍTULO 4

RESULTADOS Y DISCUSIÓN

4.1 RESULTADOS

4.1.1 SELECCIÓN DE RANGOS ÓPTIMOS

Con base en los datos recolectados, se desarrollaron varios modelos predictivos utilizando dos enfoques distintos: modelos que emplean rangos óptimos seleccionados mediante un filtrado de longitudes de onda, y modelos que utilizan el espectro completo sin aplicar dicho filtrado.

Para evaluar el rendimiento de cada modelo, se examinó su tasa de error utilizando métricas como el RMSE (Error Cuadrático Medio). Los resultados mostraron que los modelos fundamentados en rangos óptimos filtrados presentaban una tasa de error más alta, lo que sugiere que la eliminación de ciertas longitudes de onda pudo haber disminuido la capacidad del modelo para capturar información relevante de los espectros.

Por lo que se decidió de conservar toda la información espectral para la creación del modelo, evitando la pérdida de señales clave en la caracterización y cuantificación de los analitos. A partir de estos hallazgos, se decidió utilizar el espectro completo en los modelos finales para asegurar una mejor representación de la variabilidad espectral y una mayor fiabilidad en la cuantificación de ácido acético y etanol en las muestras analizadas.

4.1.2 ANÁLISIS PREEXPLORATORIO

Para el PCA se trabajó exclusivamente con los estándares “limpios” (analitos por separado), lo que permitió evaluar la distribución espectral de cada compuesto sin interferencias. En las siguientes imágenes se observa la separación clara de ambos analitos, la cual el componente 1 explica un 80.36% de la variabilidad, mientras que el componente 2 un 11,06%.

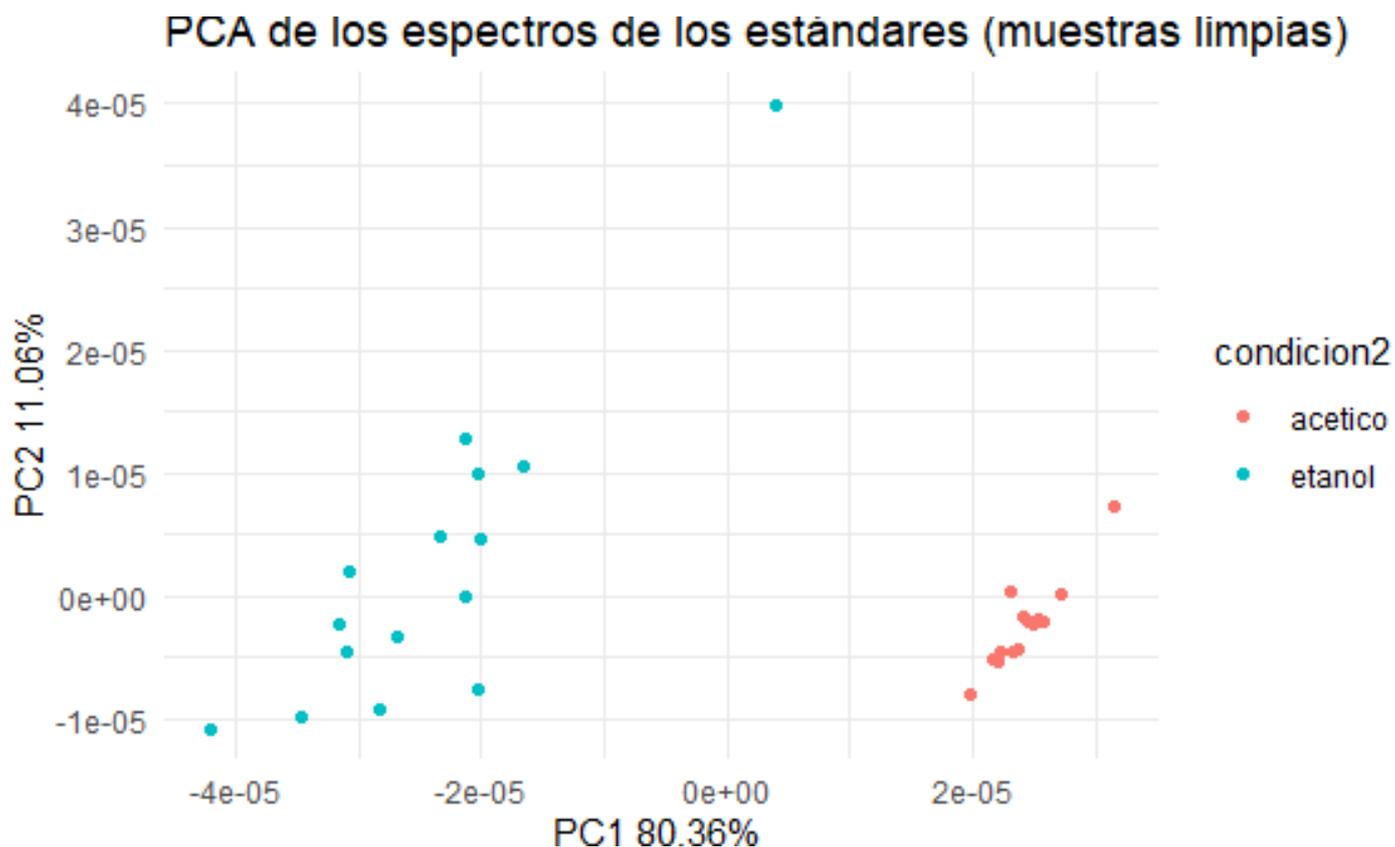


Figura 11: Gráfica de PCA de los espectros de los estándares (muestras sin contaminar)

Fuente: Propia autoría

4.1.2.1 GRÁFICO DE CONTRIBUCIÓN DE VARIABLES

A continuación, se presenta el gráfico de contribución de variables, el cual nos indica las longitudes de onda de interés. Las longitudes de onda que pasan el límite

representado por la línea roja entrecortada son consideradas de gran importancia para la identificación del etanol como del ácido acético.

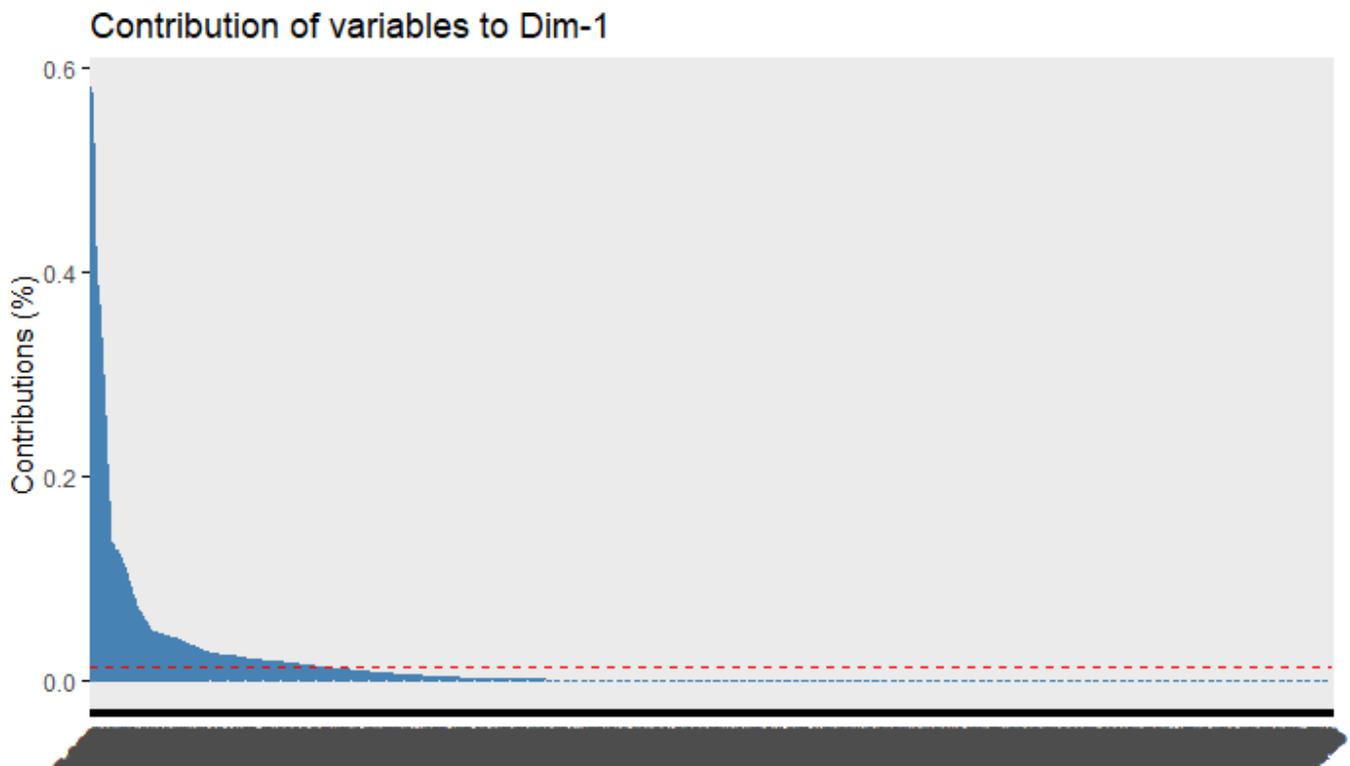


Figura 12: Gráfica de Contribución de Variables derivada del PCA

Fuente: Propia autoría

4.1.2.2 GRÁFICO DE SEDIMENTO DE COMPONENTES

En el gráfico de sedimentación de componentes, se observa que los datos procesados exponen un mayor número de dimensiones significativas. En particular, las dimensiones 1 a 3 presentan la mayor contribución a la variabilidad de los datos, lo que indica que contienen la información más relevante para el análisis posterior.

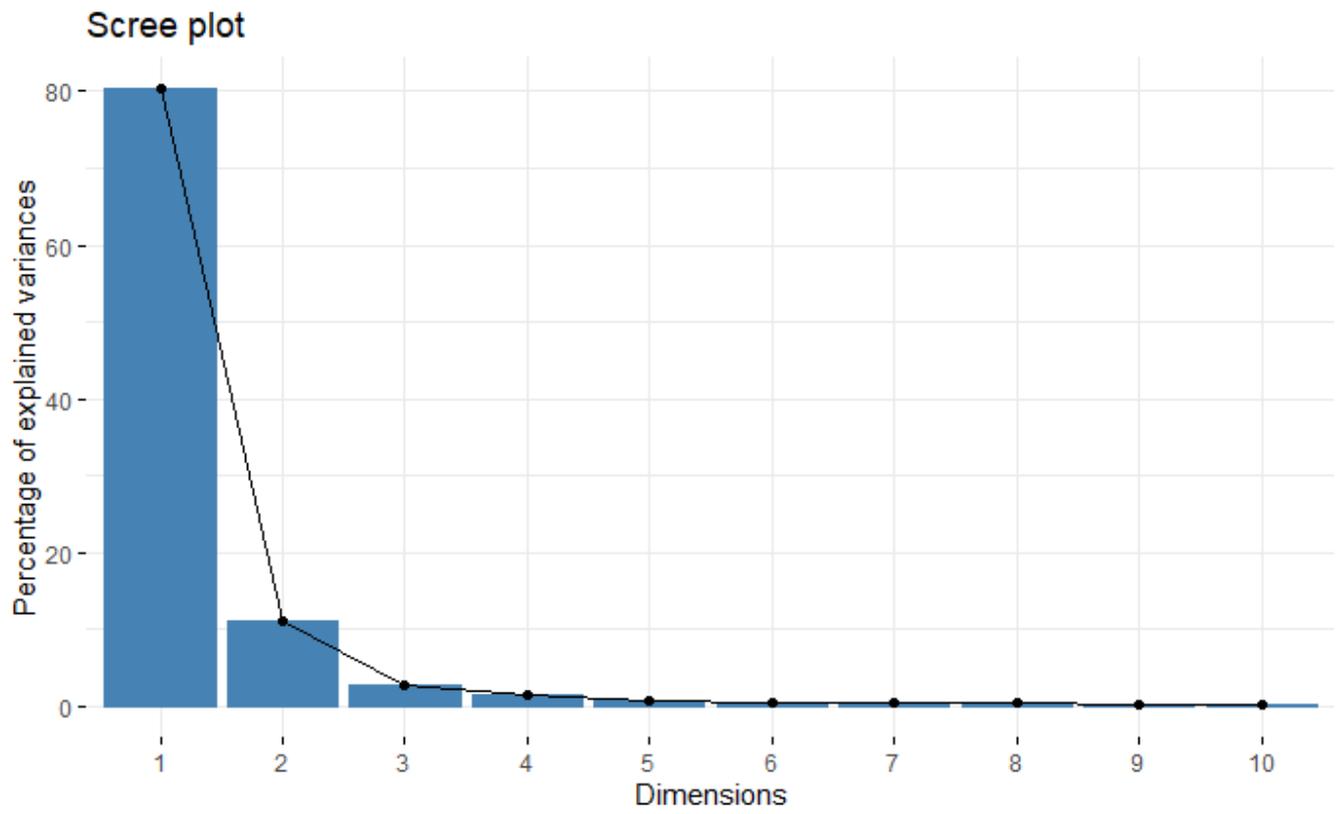


Figura 13: Gráfica de sedimentación de componentes del PCA en muestras "limpias"

Fuente: Propia autoría

4.1.3 RESULTADOS DEL ANÁLISIS PLS-DA

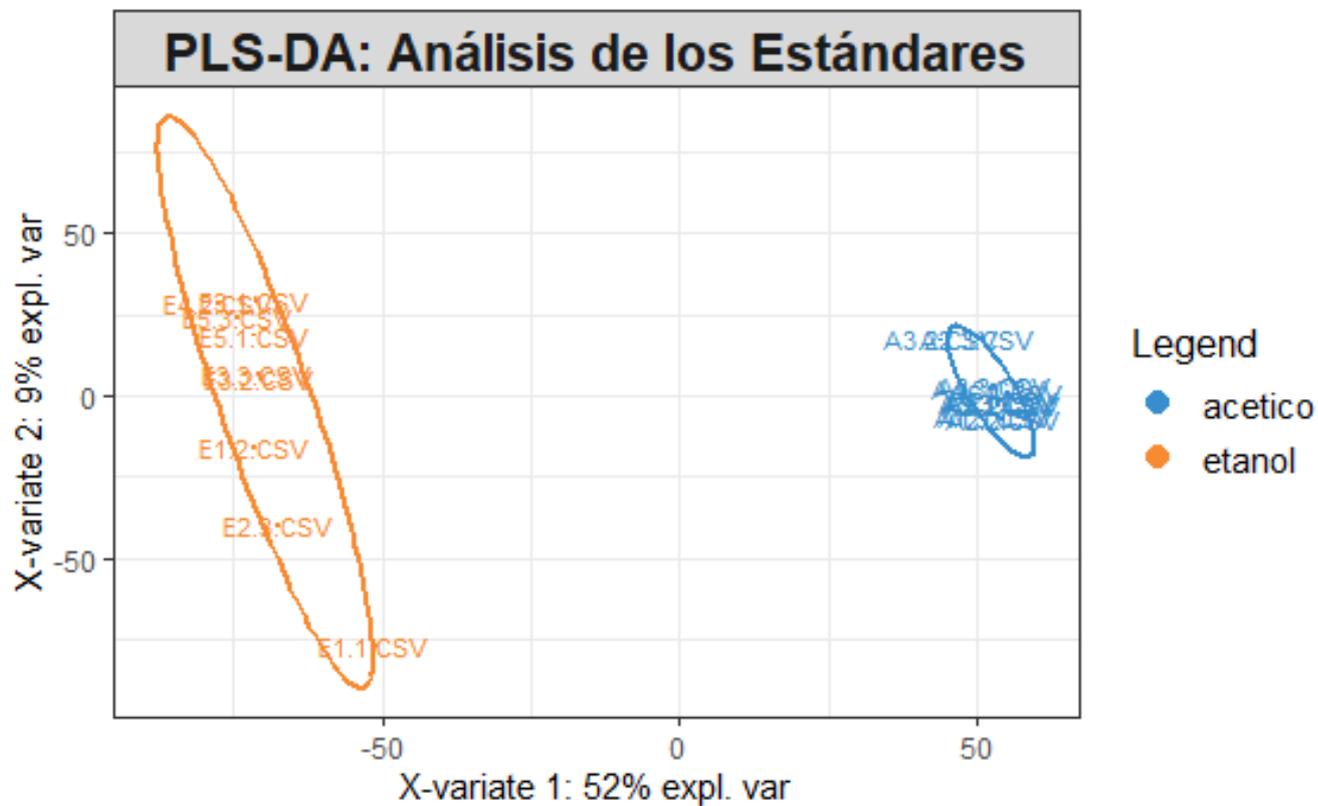


Figura 14: Gráfico de PLS-DA de estándares (analitos por separado)

Fuente: Propia autoría

Como se observa en el gráfico, los componentes están bien diferenciados, formando dos elipses correspondientes a cada grupo. La variante X1 (52%) explica la mayor variabilidad, siendo este el principal eje, mientras que la variante X2 explica un 9%. Esta separación clasifica muy bien mis analitos, confirmando su facilidad de detección.

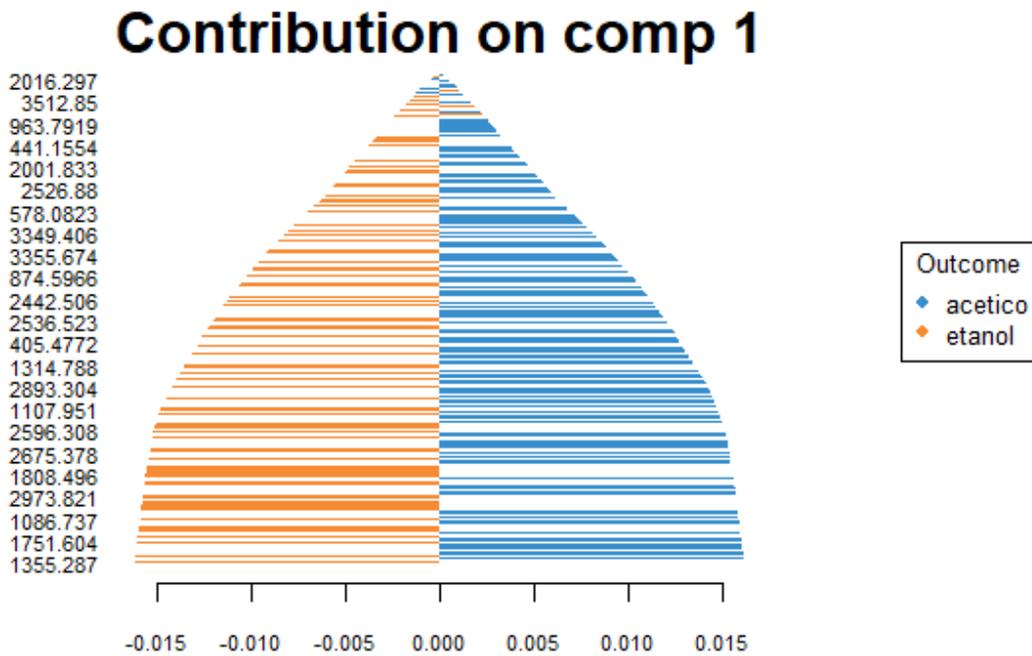


Figura 15: Gráfica de contribución del PLS-DA de mi primera componente con muestras "limpias"

Fuente: Propia autoría

El gráfico de contribución de la componente 1, representa la separación de mis analitos en un mismo eje, diferenciándolos por los colores naranja y azul, pertenecientes al ácido acético y etanol, respectivamente. En palabras sencillas, las longitudes de onda afectan la clasificación de manera opuesta dependiendo del analito, cuando una longitud de onda favorece la clasificación de ácido acético, disminuye la de etanol y viceversa.

Para evaluar el rendimiento del modelo, se utilizó la curva ROC, la cual arrojó un 100% de sensibilidad, indicando que el modelo no presentó errores en la clasificación de datos, porque son muestras puras

Este hallazgo sugiere que no habrá inconvenientes al aplicar el modelo a las muestras contaminadas controladas, ya que la clasificación inicial de los estándares se realizó sin ambigüedad.

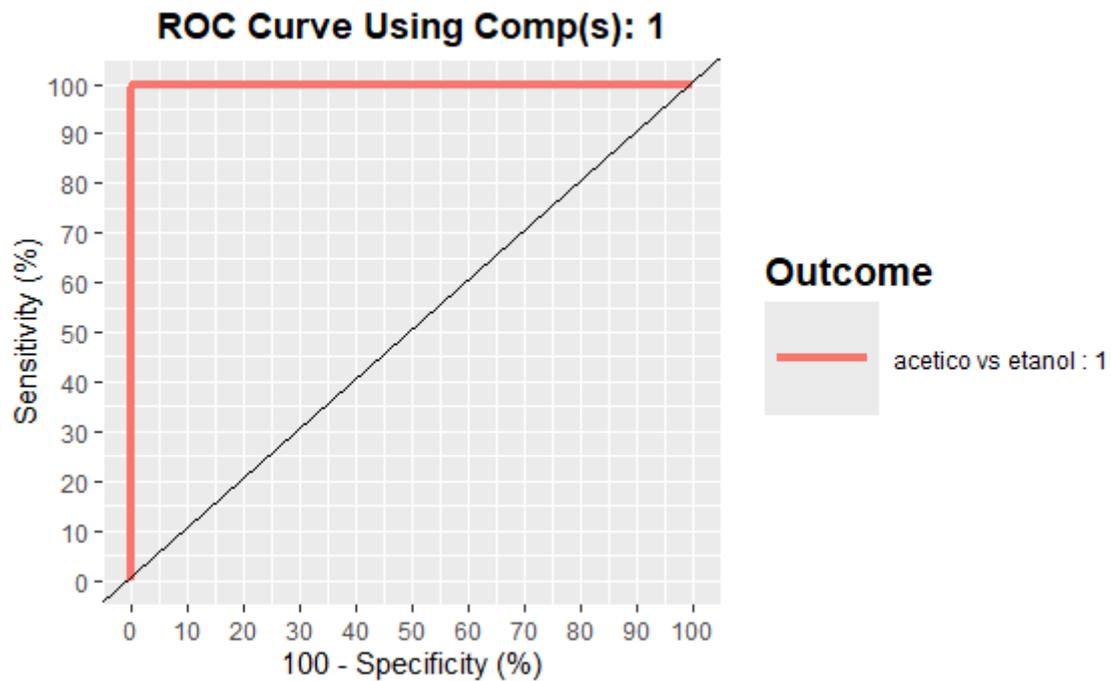


Figura 16: Curva de ROC para el análisis de PLS-DA

Fuente: Propia autoría

4.1.4 RESULTADOS DEL ANÁLISIS sPLS

Se resume el modelo efectuado en R de la siguiente manera:

- **Tipo de modelo:** Regresión
- **Número de componentes:** 2
- **Dimensión de los datos:**
 - **X (predictores):** 45 x 7469 (45 muestras, 7469 variables espectrales)
 - **Y (variable respuesta):** 45 x 2 (45 muestras, 2 variables respuesta)
- Variables seleccionadas en cada componente (keepX y keepY):

- **keepX:** 2200 y 3500 variables retenidas en cada componente
- **keepY:** 1 y 2 variables retenidas en cada componente

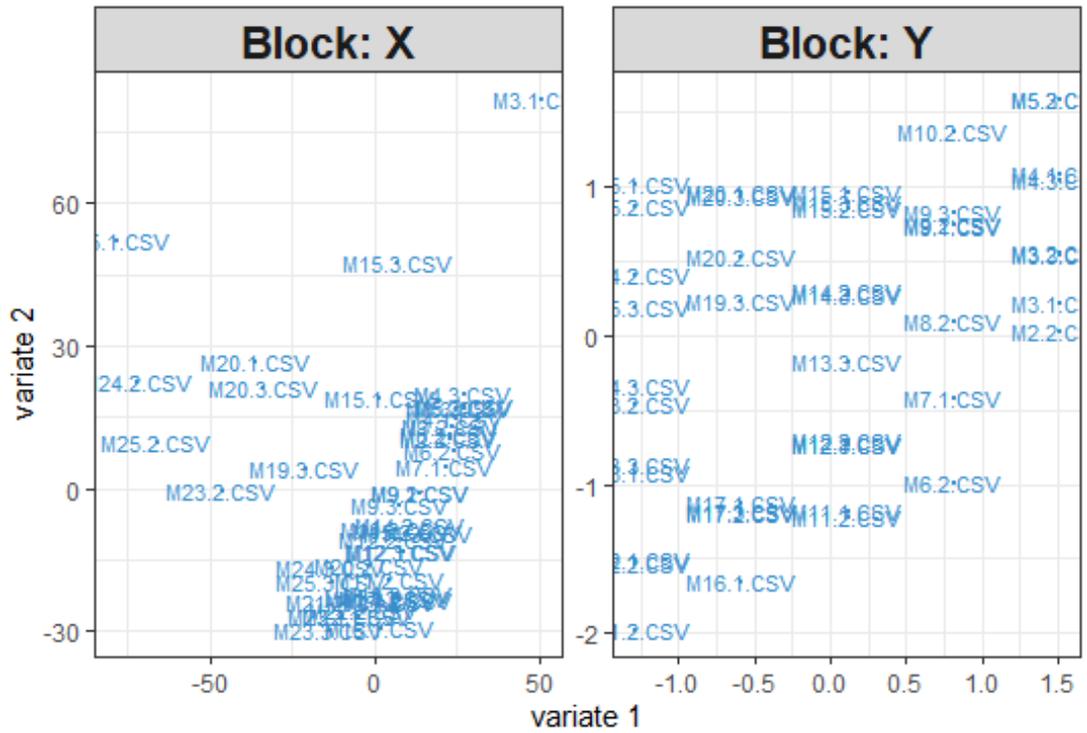


Figura 17: Gráfica de bloques(X,Y) de los componentes principales del PLS

Fuente: Propia autoría

Como podemos observar en el gráfico, hay dos bloques uno correspondiente a X y otro a Y, estos compartimientos representan las variables predictoras y la otra, la variable de respuesta, respectivamente. Por medio del gráfico confirmamos que el modelo está identificando patrones significativos entre los datos, mostrando así las relaciones claves de ambas matrices.

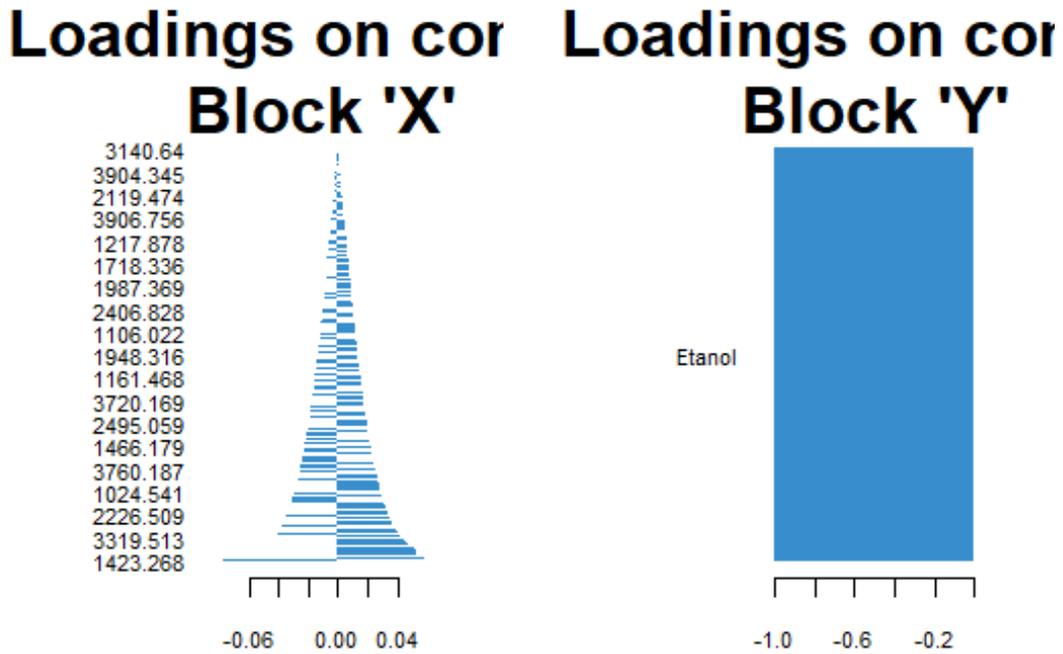


Figura 18: Gráfico de contribución de cada bloque (X,Y) del PLS

Fuente: Propia autoría

En el gráfico de contribución de variables de los bloques X y Y, nos indica que ciertos espectros pueden predecir muy bien la concentración de etanol, ya que el bloque Y corresponde a la variable de respuesta, demostrando que mi bloque X, contribuye de manera similar en su predicción.

4.1.4.1 RMSE DEL MODELO

Los resultados de mi tasa de error, se la puede evidenciar en la siguiente tabla:

COMPONENTE	ANALITO	RMSE
Componente 1	Etanol	0.11772782
Componente 1	Ácido Acético	0.34047798
Componente 2	Etanol	0.08451969
Componente 2	Ácido Acético	0.33542750

Tabla 2: Tabla de los resultados del RMSE calculados del modelo (PLS)

Fuente: Propia autoría

Conociendo estos resultados, podemos decir que el modelo generado es confiable ante sus predicciones, ya que las tasas de error son bajas.

4.1.5 PREDICCIÓN DE MUESTRAS DE VINAGRE

Al aplicar el modelo PLS a las muestras de vinagre, se obtuvieron los resultados de concentraciones de las muestras analizadas de cada analito, la cual se mostrará en la tabla, la cual la columna cuyo nombre es “madre” indica la muestra casera:

COMPONENTE	ANALITO	MADRE	COMERCIAL
Componente 1	Etanol	0.7845601	0.7615159
Componente 1	Ácido Acético	5.5873886	5.5727367
Componente 2	Etanol	0.7679627	0.8078596
Componente 2	Ácido Acético	5.6222875	5.4752915

Tabla 3: Tabla de resultados de predicción de las muestras de vinagre, los resultados se encuentran en % (v/v)

Fuente: Propia autoría

Es importante resaltar que los componentes 1 y 2, son los que representan la afinidad que posee el modelo por cada analito; como observamos en la Figura 18, el bloque Y analiza las predicciones de etanol, por lo que se indica que el bloque Y corresponde a la componente 2.

Debido a que la componente 2 tiene una mejor relación con la predicción del etanol, se espera una detección más precisa del etanol en comparación el componente 1 que se encuentra influenciado por el ácido acético.

Debido a que el modelo presentó un porcentaje bajo de error, concluimos que las respuestas generadas son confiables.

En el ANEXO 2:ANEXO 1:

Tabla Completa del Experimento

AcAcetico	Etanol	Agua	Codigo
5	0,5	0,945	M1
5,25	0,5	0,9425	M2
5,5	0,5	0,94	M3
5,75	0,5	0,9375	M4
6	0,5	0,935	M5
5	0,625	0,94375	M6
5,25	0,625	0,94125	M7
5,5	0,625	0,93875	M8
5,75	0,625	0,93625	M9
6	0,625	0,93375	M10
5	0,75	0,9425	M11
5,25	0,75	0,94	M12
5,5	0,75	0,9375	M13
5,75	0,75	0,935	M14
6	0,75	0,9325	M15
5	0,875	0,94125	M16
5,25	0,875	0,93875	M17
5,5	0,875	0,93625	M18
5,75	0,875	0,93375	M19
6	0,875	0,93125	M20
5	1	0,94	M21
5,25	1	0,9375	M22
5,5	1	0,935	M23
5,75	1	0,9325	M24
6	1	0,93	M25

Fuente: Propia Autoría

ANEXO 2:

Tabla de Comparación de Longitudes de Onda entre los Estándares (Analitos: Ác Acético y Etanol por separado) y las Muestras Controladas Contaminadas, se presenta una tabla de las longitudes de onda de interés la cual nos indica qué número de onda pertenece al Etanol o Ác Acético respectivamente, tanto en las muestras contaminadas controladas como en las muestras “limpia”. Esta tabla proporciona una última verificación sobre la importancia de trabajar con el espectro completo, evitando la eliminación de información relevante mediante filtrados.

Además, permite analizar el desplazamiento de onda de los analitos, lo que contribuye a una mejor comprensión de sus variaciones espectrales en las muestras contaminadas.

4.2 DISCUSIÓN

4.2.1 DETERMINACIÓN DE RANGOS ESPECTRALES ÓPTIMOS

Para el preprocesamiento de datos espectrales, se evaluaron diferentes métodos con el fin de seleccionar el método óptimo. Entre los procesos aplicados se encuentran: corrección de línea base (métodos Rubberband y ALS), normalización de áreas, suavizado con el filtro Savitzky-Golay, primera y segunda derivada y finalmente escalamiento Z-score. La elección del método de corrección de línea base jugó un papel importante en la obtención de los espectros y su análisis.

El método Rubberband ajusta una envolvente convexa a los puntos más bajos del espectro, permitiendo corregir la inclinación de la línea base sin afectar la forma de los picos(Wang et al., 2018). A diferencia del método ALS, el cual mostró mejores resultados, permitiendo la eliminación de toda señal de interferencia no deseada(Peng et al., 2010), eliminando así el pico característico del enlace de -OH (agua).

Se escogió el método ALS, debido a que se busca resaltar los enlaces predominantes de los analitos, así se evita mayor variabilidad en el modelo al mantener el pico de agua.

En cuestión del normalizado por área y su suavizado, no hubo inconvenientes u observaciones significativas, sin embargo, para la primera y segunda derivada, el cual se ocupó el filtro de Savitzky-Golay, se observaron diferencias significativas, es importante

mencionar que los datos tratados por el método de Rubberband fueron sometidos a la primera derivada, mientras que los datos tratados por ALS se les realizó directamente la 2da derivada. A ambas configuraciones se les realizó un escalado de Z score.

Después de establecer la configuración ideal para el preprocesamiento, se optó por trabajar con el espectro completo, dado que los resultados obtenidos con esta configuración superaron en precisión y estabilidad al modelo. A diferencia de los intentos anteriores de filtrado de variables, utilizar el espectro sin restricciones permitió captar toda la información relevante de los analitos y evitar la pérdida de longitudes de onda que podrían ser significativas. Esta estrategia mejoró la capacidad predictiva del modelo y redujo la pérdida de información espectral importante.

4.2.2 INTERPRETACIÓN DE MODELOS PREDICTIVOS MULTIDIMENSIONALES

El análisis mediante PCA (Análisis de Componentes Principales) facilitó la evaluación de la variabilidad espectral contenida en los datos y la exploración de la separación entre los analitos de interés. Se determinó que la primera componente principal (PC1) explicaba el 80.36% de la variabilidad, lo que sugiere que los analitos presentan diferencias espectrales significativas, permitiendo, así, su diferenciación precisa. Esta clara distinción avaló la posibilidad de aplicar métodos de clasificación y selección de variables para desarrollar modelos predictivos.

Con base en los resultados del PCA, se elaboró una gráfica de contribución de variables con el fin de identificar las longitudes de onda más significativas. En un primer momento, se filtraron aquellas que tenían una contribución superior al 0.01%. No obstante, se decidió trabajar con el espectro completo preprocesado, para evitar la pérdida de longitudes de onda que podrían ser relevantes para la modelización. Esta elección garantizó la conservación de la mayor cantidad de información espectral posible y evitó la omisión de señales potencialmente significativas para la cuantificación de los analitos.

Para clasificar los espectros, se utilizó el método de Análisis Discriminante por Mínimos Cuadrados Parciales (PLSDA). Este método es más eficiente en la identificación de las longitudes de onda más pertinentes para la diferenciación de los analitos.

La evaluación del modelo se llevó a cabo mediante la curva ROC (Receiver Operating Characteristic), cuyas mediciones mostraron una sensibilidad del 100% (Figura

16). Aunque en otros contextos estadísticos un valor tan elevado podría indicar sobreajuste, en este caso, el resultado es coherente con la naturaleza de los datos, ya que se trabajó con muestras puras y libres de contaminación. Esto confirma que los espectros analizados presentan diferencias espectrales marcadas que permiten una clasificación confiable.

Finalmente, para construir el modelo predictivo, se eligió el método sPLS (Mínimos Cuadrados Parciales Escasos) en lugar del PLS tradicional. La principal razón para esta elección fue su habilidad para la selección de variables, lo que permitió crear un modelo más eficiente y robusto al disminuir la redundancia espectral sin afectar la precisión en la predicción de las concentraciones de los analitos.

4.2.3 EVALUACIÓN DEL MODELO PLS Y PREDICCIÓN EN MUESTRAS DE VINAGRE

Con respecto a las muestras de vinagre de manzana, el modelo indicó concentraciones de ácido acético y etanol que estaban dentro de los límites establecidos por el Codex Stan 162-1987. Esto confirma la capacidad predictiva del modelo, así como su aplicabilidad en matrices reales.

Como se puede notar fácilmente, los datos ya son buenos, pero se pueden mejorar calibrando el modelo con más muestras para que se pueda manejar una mayor variabilidad.

4.2.4 COMPARACIÓN CON ESTUDIOS PREVIOS Y LIMITACIONES DEL MODELO

Muchos estudios han investigado extensamente la aplicación de FTIR combinado con modelos quimiométricos para la identificación y cuantificación de compuestos disueltos en matrices líquidas y sólidas.

En el presente estudio, la aplicación de PLS-DA y PLS también permitió distinguir y cuantificar el ácido acético y el etanol en muestras de vinagre, cuyos resultados coinciden con estudios previos que adoptaron este enfoque.

Por ejemplo, investigaciones como la de Rahmania et al. (2015) han demostrado la efectividad del PCA y PLS para la identificación de adulteraciones en productos alimentarios, específicamente en el análisis de carne de rata en albóndigas. Asimismo,

estudios como el de Smok-Kalwat et al. (2024) evidencian la capacidad de FTIR en combinación con machine learning para la detección de biomarcadores en biopsias líquidas, aplicando técnicas similares a las utilizadas en este trabajo, como PCA y regresión logística.

En el campo de la industria alimentaria, investigaciones como las de Sahlan et al. (2019) y Vinciguerra et al. (2019) han utilizado ATR-FTIR para la detección de adulteraciones en miel y leche, respectivamente, empleando modelos predictivos para la autenticación de productos. En comparación con estos estudios, la presente investigación refuerza la aplicación de FTIR como una herramienta rápida y eficiente, demostrando su viabilidad en la caracterización de matrices fermentadas como el vinagre.

4.2.4.1 Limitantes

A pesar de los resultados prometedores obtenidos en este estudio, se identifican algunas limitaciones que podrían influir en la precisión y aplicabilidad del modelo:

- Presencia de otros compuestos en las muestras de vinagre de manzana
- Falta cotejamiento de los números de onda con bases de datos o muestras puras, como recomendación o limitantes.

CAPÍTULO 5

5.1 CONCLUSIONES

Después de analizar los resultados concluyo que:

- Utilizar un rango espectral específico no era la alternativa más adecuada para modelar los analitos de interés. Por lo tanto, se optó por trabajar con el espectro completo, lo cual facilitó una mayor precisión en el modelo y disminuyó la tasa de error al predecir las concentraciones de ácido acético y etanol en el vinagre.
- Se desarrolló un modelo utilizando sPLS (Sparse Partial Least Squares), que demostró un rendimiento excelente en la predicción simultánea de los dos analitos. Optar por este método favoreció la maximización de la relevancia de las variables espectrales y optimizó la capacidad predictiva del modelo.

- Los resultados de la evaluación del modelo utilizando muestras de vinagre estuvieron dentro de los parámetros previamente definidos. Esto confirma que el modelo es capaz de cuantificar de manera precisa tanto el ácido acético como el etanol, reforzando la teoría de que hay relaciones funcionales entre las longitudes de onda y las concentraciones de los analitos analizados.
- Los resultados obtenidos respaldan la viabilidad de crear modelos de regresión multidimensional que permitan la predicción simultánea de ácido acético y etanol en el vinagre, lo que podría facilitar su uso en la comprobación de los límites legales en productos comerciales.

De esta forma, los resultados obtenidos respaldan la hipótesis formulada, al evidenciar una relación funcional entre las longitudes de onda y las concentraciones de los analitos, lo que facilitó el éxito en el desarrollo del modelo predictivo. Además, se responde a la pregunta de investigación, confirmando que es factible crear modelos de regresión capaces de predecir de manera simultánea los niveles de ácido acético y etanol en el vinagre.

5.2 RECOMENDACIONES

- Incluir mayor número de muestras de vinagre de manzana, tanto artesanales/caseras como comerciales para mejorar la generalización del modelo.
- Explorar otros métodos de preprocesado, ya sea en la corrección de línea base, normalización o derivada.
- Contrastar los resultados obtenidos del proyecto con métodos tradicionales de cuantificación como puede ser la cromatografía de gases o HPLC, para validar la precisión del modelo.
- Extender la aplicación del modelo a diferentes productos de preferencia a productos que puedan presentar los analitos estudiados, es decir, contengan la presencia de ácido acético y etanol.
- Evaluar la viabilidad de implementar este modelo en laboratorios de control de calidad.

DECLARACIÓN DE DISPONIBILIDAD DEL CÓDIGO

Los scripts y datos utilizados en este estudio están disponibles en el siguiente repositorio:

<https://github.com/gabyg108/TESIS.git>

Se recomienda revisar los archivos script.R bajo el nombre de TESIS FINAL para replicar el análisis.

REFERENCIAS BIBLIOGRÁFICAS

Athavale, S. V., Simon, A., Houk, K. N., & Denmark, S. E. (2020). Demystifying the asymmetry-amplifying, autocatalytic behaviour of the Soai reaction through structural, mechanistic and computational studies. *Nature Chemistry*, *12*, 412–423. <https://doi.org/10.1038/s41557-020-0421-8>

Becerra, M., & Goos, P. (2021). Bayesian I-optimal designs for choice experiments with mixtures. *Chemometrics and Intelligent Laboratory Systems*, *217*. <https://doi.org/10.1016/j.chemolab.2021.104395>

Berthomieu, C., & Hienerwadel, R. (2009). Fourier transform infrared (FTIR) spectroscopy. In *Photosynthesis Research* (Vol. 101, pp. 157–170). <https://doi.org/10.1007/s11120-009-9439-x>

CAPITULO II Espectroscopia del infrarrojo 2.1 Región del infrarrojo. (n.d.).

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>

Codex stan 162-1987 norma del codex para el vinagre. (n.d.).

Devianti, D., Sufardi, S., Zulfahrizal, Z., & Munawar, A. A. (2019). Rapid and Simultaneous Detection of Hazardous Heavy Metals Contamination in Agricultural Soil Using Infrared Reflectance Spectroscopy. *IOP Conference Series: Materials Science and Engineering*, *506*(1). <https://doi.org/10.1088/1757-899X/506/1/012008>
document. (n.d.).

Fisher, R. A. (1936). The use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, *7*, 179–188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>

FTIR | FTIR Spectroscopy Academy | Thermo Fisher Scientific - IE. (n.d.). <https://www.thermofisher.com/ie/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fourier-transform-infrared-spectroscopy/resources/ftir-spectroscopy-academy.html>

- Gutiérrez Pulido, H., & de la Vara Salazar, R. (n.d.). *Análisis y diseño de experimentos*.
www.FreeLibros.org
- Henning Teickner. (2018). *Introduction to the 'ir' package*. [https://Cran.r-Project.Org/Web/Packages/Ir/Vignettes/Ir-Introduction.Html](https://cran.r-project.org/web/packages/ir/vignettes/ir-introduction.html).
- Hodgkins, S. B., Richardson, C. J., Dommmain, R., Wang, H., Glaser, P. H., Verbeke, B., Winkler, B. R., Cobb, A. R., Rich, V. I., Missilmani, M., Flanagan, N., Ho, M., Hoyt, A. M., Harvey, C. F., Vining, S. R., Hough, M. A., Moore, T. R., Richard, P. J. H., De La Cruz, F. B., ... Chanton, J. P. (2018). Tropical peatland carbon storage linked to global latitudinal trends in peat recalcitrance. *Nature Communications*, 9. <https://doi.org/10.1038/s41467-018-06050-2>
- 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. <https://doi.org/10.1186/1471-2105-12-253>
- Lizarazo, S., Villamizar, J., Petit, V., Orozco, O., & Rueda, A. (2022, June 8). *Código PCA*. [https://Rpubs.Com/Andress20/Codigo_pca](https://rpubs.com/Andress20/Codigo_pca).
- Lorenzo, J. (n.d.). *VISIÓN GENERAL Y OBJETIVOS*. <https://ansenuza.unc.edu.ar/comunidades/handle/11086.1/1202>
- Luo, X., Takahashi, T., Kyo, K., & Zhang, S. (2012). Wavelength selection in vis/NIR spectra for detection of bruises on apples by ROC analysis. *Journal of Food Engineering*, 109(3), 457–466. <https://doi.org/10.1016/j.jfoodeng.2011.10.035>
- Movasaghi, Z., Rehman, S., & Rehman, I. U. (2008). Fourier transform infrared (FTIR) spectroscopy of biological tissues. In *Applied Spectroscopy Reviews* (Vol. 43, pp. 134–179). <https://doi.org/10.1080/05704920701829043>
- Nguyen, D. V, & Rocke, D. M. (2002). Tumor classification by partial least squares using microarray gene expression data. In *BIOINFORMATICS* (Vol. 18, Issue 1).
- 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-Enciso, M., & Tenenhaus, M. (2003a). Prediction of clinical outcome with microarray data: A partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics*, *112*, 581–592. <https://doi.org/10.1007/s00439-003-0921-9>
- Pérez-Enciso, M., & Tenenhaus, M. (2003b). Prediction of clinical outcome with microarray data: A partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics*, *112*(5–6), 581–592. <https://doi.org/10.1007/s00439-003-0921-9>
- ¿Qué es el análisis de componentes principales (PCA)? | IBM. (n.d.). <https://www.ibm.com/es-es/topics/principal-component-analysis>.
- Rahmania, H., Sudjadi, & Rohman, A. (2015). The employment of FTIR spectroscopy in combination with chemometrics for analysis of rat meat in meatball formulation. *Meat Science*, *100*, 301–305. <https://doi.org/10.1016/j.meatsci.2014.10.028>
- 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, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*, 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Smok-Kalwat, J., Gózdź, S., Macek, P., Kalwat, Z., Khalavka, M., Rząd, W., Stepulak, A., & Depciuch, J. (2024). Serum and plasma as a good candidates of body fluids for detection lung cancer by FTIR liquid biopsy. *Scientific Reports*, *14*. <https://doi.org/10.1038/s41598-024-81649-8>
- Two methods for baseline correction of spectral data • NIRPY Research*. (n.d.). <https://nirpyresearch.com/two-methods-baseline-correction-spectral-data/>.
- Vinciguerra, L. L., Marcelo, M. C. A., Motta, T. M. C., Meneghini, L. Z., Bergold, A. M., & Ferrão, M. F. (2019). Chemometric tools and FTIR-ATR spectroscopy applied in milk adulterated with cheese whey. *Química Nova*, *42*(3), 249–254. <https://doi.org/10.21577/0100-4042.20170337>

Wang, X., Lü, S. L., Li, Y., Wei, H. Y., & Chen, X. (2018). Automatic Baseline Correction of Gas Spectra Based on Baseline Drift Model. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis*, 38, 3946–3951. [https://doi.org/10.3964/j.issn.1000-0593\(2018\)12-3946-06](https://doi.org/10.3964/j.issn.1000-0593(2018)12-3946-06)

ANEXOS

ANEXO 1:

Tabla Completa del Experimento

AcAcetico	Etanol	Agua	Codigo
5	0,5	0,945	M1
5,25	0,5	0,9425	M2
5,5	0,5	0,94	M3
5,75	0,5	0,9375	M4
6	0,5	0,935	M5
5	0,625	0,94375	M6
5,25	0,625	0,94125	M7
5,5	0,625	0,93875	M8
5,75	0,625	0,93625	M9
6	0,625	0,93375	M10
5	0,75	0,9425	M11
5,25	0,75	0,94	M12
5,5	0,75	0,9375	M13
5,75	0,75	0,935	M14
6	0,75	0,9325	M15
5	0,875	0,94125	M16
5,25	0,875	0,93875	M17
5,5	0,875	0,93625	M18
5,75	0,875	0,93375	M19
6	0,875	0,93125	M20
5	1	0,94	M21
5,25	1	0,9375	M22
5,5	1	0,935	M23
5,75	1	0,9325	M24
6	1	0,93	M25

Fuente: Propia Autoría

ANEXO 2:

Tabla de Comparación de Longitudes de Onda entre los Estándares (Analitos: Ác Acético y Etanol por separado) y las Muestras Controladas Contaminadas

Onda	Contaminado	Limpio
516.8509	AC	
517.333	AC	
517.8152	AC	
558.7968	AC	AC
559.2789	AC	AC
559.7611	AC	AC
560.2432	AC	AC
560.7253	AC	AC
625.3317	AC	AC
631.5995	AC	AC
632.0816	AC	AC
846.6326	AC	ET
847.1147	AC	ET
847.5969	AC	ET
848.079	AC	ET
865.9181	AC	AC
866.4002	AC	AC
866.8823	AC	AC
873.1501	AC	
873.6323	AC	
874.1144	AC	
874.5966	AC	
875.0787	AC	
884.7214	AC	ET
885.2036	AC	ET
885.6857	AC	ET
886.1678	AC	ET
886.65	AC	ET
887.1321	AC	ET
887.6143	AC	ET
888.0964	AC	ET
888.5785	AC	ET
889.0607	AC	ET
889.5428	AC	ET
900.632	AC	AC
901.1141	AC	AC

901.5962	AC	AC
902.0784	AC	AC
902.5605	AC	AC
903.0426	AC	AC
903.5248	AC	AC
904.0069	AC	AC
904.489	AC	AC
925.7031	AC	ET
926.1852	AC	ET
926.6673	AC	ET
927.1495	AC	ET
927.6316	AC	ET
928.1137	AC	ET
928.5959	AC	ET
929.078	AC	ET
932.4529	AC	AC
932.9351	AC	AC
933.4172	AC	AC
933.8994	AC	AC
998.0236	AC	ET
998.5057	AC	ET
998.9879	AC	ET
999.47	AC	ET
1007.666	AC	ET
1008.148	AC	ET
1008.631	AC	ET
1009.113	AC	ET
1009.595	AC	ET
1010.077	AC	ET
1010.559	AC	ET
1011.041	AC	ET
1011.523	AC	ET
1012.006	AC	ET
1012.488	AC	ET
1012.97	AC	ET
1013.452	AC	ET
1013.934	AC	ET
1014.416	AC	ET
1014.898	AC	ET
1015.381	AC	ET
1015.863	AC	
1016.345	AC	
1016.827	AC	AC
1021.648	AC	AC
1022.13	AC	AC
1022.613	AC	AC

1023.095	AC	AC
1023.577	AC	AC
1024.059	AC	AC
1024.541	AC	AC
1025.023	AC	AC
1025.505	AC	AC
1025.988	AC	AC
1026.47	AC	AC
1026.952	AC	AC
1027.434	AC	AC
1027.916	AC	AC
1028.398	AC	AC
1028.88	AC	AC
1029.363	AC	AC
1029.845	AC	AC
1030.327	AC	AC
1030.809	AC	AC
1031.291	AC	AC
1031.773	AC	AC
1032.255	AC	AC
1032.738	AC	AC
1033.22	AC	AC
1033.702	AC	AC
1041.416	AC	AC
1041.898	AC	AC
1042.38	AC	AC
1042.862	AC	AC
1043.344	AC	AC
1043.827	AC	AC
1047.202	AC	ET
1047.684	AC	ET
1048.166	AC	ET
1048.648	AC	ET
1049.13	AC	ET
1049.612	AC	ET
1050.094	AC	ET
1050.577	AC	ET
1051.059	AC	ET
1051.541	AC	ET
1052.023	AC	ET
1052.505	AC	ET
1052.987	AC	ET
1053.469	AC	ET
1066.487	AC	AC
1066.969	AC	AC
1067.451	AC	AC

1067.933	AC	AC
1068.416	AC	AC
1068.898	AC	AC
1079.987	AC	AC
1080.469	AC	AC
1080.951	AC	AC
1081.433	AC	AC
1081.916	AC	AC
1082.398	AC	AC
1082.88	AC	AC
1083.362	AC	AC
1083.844	AC	AC
1084.326	AC	AC
1084.808	AC	AC
1085.29	AC	AC
1085.772	AC	ET
1173.521	AC	
1174.004	AC	AC
1174.486	AC	AC
1174.968	AC	AC
1225.11	AC	ET
1225.592	AC	ET
1226.074	AC	ET
1226.557	AC	ET
1227.039	AC	ET
1227.521	AC	ET
1231.86	AC	ET
1232.342	AC	ET
1254.521	AC	ET
1255.003	AC	ET
1255.485	AC	ET
1255.967	AC	ET
1256.449	AC	
1335.52	AC	AC
1341.787	AC	ET
1342.269	AC	ET
1342.751	AC	ET
1343.234	AC	ET
1343.716	AC	ET
1344.198	AC	ET
1344.68	AC	ET
1366.858	AC	AC
1367.341	AC	AC
1367.823	AC	AC
1368.305	AC	AC
1368.787	AC	AC

1369.269	AC	AC
1369.751	AC	AC
1370.233	AC	AC
1370.716	AC	AC
1371.198	AC	AC
1371.68	AC	AC
1372.162	AC	AC
1372.644	AC	AC
1373.126	AC	AC
1373.608	AC	AC
1374.09	AC	AC
1374.573	AC	AC
1375.055	AC	AC
1375.537	AC	AC
1376.019	AC	AC
1376.501	AC	AC
1376.983	AC	AC
1377.465	AC	AC
1377.948	AC	AC
1378.43	AC	AC
1378.912	AC	AC
1379.394	AC	AC
1381.323	AC	AC
1381.805	AC	AC
1382.287	AC	AC
1382.769	AC	AC
1383.251	AC	AC
1383.733	AC	AC
1384.215	AC	AC
1384.698	AC	AC
1385.18	AC	AC
1385.662	AC	AC
1386.144	AC	
1386.626	AC	
1387.108	AC	
1387.59	AC	
1388.073	AC	
1388.555	AC	ET
1389.037	AC	ET
1389.519	AC	ET
1390.001	AC	ET
1390.483	AC	ET
1390.965	AC	ET
1391.447	AC	ET
1391.929	AC	ET
1392.412	AC	ET

1392.894	AC	ET
1393.376	AC	ET
1393.858	AC	ET
1394.34	AC	ET
1394.822	AC	ET
1395.304	AC	ET
1395.787	AC	ET
1396.269	AC	ET
1396.751	AC	ET
1397.233	AC	ET
1397.715	AC	ET
1399.162	AC	
1415.554	AC	AC
1416.036	AC	AC
1416.519	AC	AC
1420.858	AC	ET
1421.34	AC	ET
1421.822	AC	ET
1422.304	AC	ET
1422.786	AC	ET
1423.268	AC	ET
1423.75	AC	ET
1424.233	AC	ET
1440.625	AC	AC
1441.107	AC	AC
1441.59	AC	AC
1442.072	AC	AC
1442.554	AC	AC
1443.036	AC	AC
1443.518	AC	AC
1444	AC	AC
1444.482	AC	AC
1444.965	AC	AC
1445.447	AC	AC
1445.929	AC	AC
1446.411	AC	AC
1446.893	AC	AC
1447.375	AC	AC
1447.857	AC	AC
1448.34	AC	AC
1448.822	AC	AC
1449.304	AC	AC
1449.786	AC	AC
1450.268	AC	AC
1450.75	AC	AC
1454.607	AC	AC

1455.089	AC	AC
1455.572	AC	AC
1456.054	AC	
1456.536	AC	
1461.357	AC	ET
1461.839	AC	ET
1465.697	AC	ET
1466.179	AC	ET
1466.661	AC	ET
1467.143	AC	ET
1467.625	AC	ET
1468.107	AC	ET
1468.589	AC	ET
1469.072	AC	ET
1469.553	AC	ET
1470.036	AC	ET
1470.518	AC	ET
1474.857	AC	ET
1475.339	AC	ET
1475.821	AC	ET
1476.303	AC	ET
1476.786	AC	
1477.268	AC	
1481.125	AC	AC
1481.607	AC	
1482.089	AC	
1484.982	AC	AC
1485.464	AC	AC
1485.946	AC	AC
1486.428	AC	AC
1486.91	AC	
1491.732	AC	
1529.821	AC	AC
1530.303	AC	AC
1530.785	AC	AC
1531.267	AC	AC
1531.749	AC	AC
1533.196	AC	AC
1533.678	AC	AC
1534.16	AC	AC
1534.642	AC	AC
1535.124	AC	AC
1535.606	AC	AC
1539.946	AC	AC
1540.428	AC	AC
1548.142	AC	AC

1548.624	AC	
1549.106	AC	
1549.588	AC	
1550.071	AC	
1550.552	AC	
1559.713	AC	ET
1573.695	AC	ET
1574.177	AC	ET
1574.659	AC	ET
1575.142	AC	ET
1575.624	AC	ET
1576.106	AC	ET
1576.588	AC	ET
1578.517	AC	ET
1578.999	AC	ET
1579.481	AC	ET
1579.963	AC	ET
1580.445	AC	ET
1580.927	AC	ET
1581.409	AC	ET
1581.892	AC	ET
1582.374	AC	ET
1582.856	AC	ET
1583.338	AC	ET
1583.82	AC	ET
1584.302	AC	ET
1738.104	AC	AC
1738.586	AC	AC
1739.068	AC	AC
1739.55	AC	AC
1774.746	AC	AC
1780.05	AC	AC
1780.532	AC	AC
1781.014	AC	AC
1781.496	AC	AC
1781.978	AC	AC
1804.639	AC	ET
1805.121	AC	ET
1805.603	AC	ET
1806.085	AC	ET
1806.567	AC	ET
1807.05	AC	ET
1825.371	AC	AC
1825.853	AC	AC
1861.049	AC	AC
1861.531	AC	AC

1862.013	AC	AC
1862.495	AC	AC
1880.816	AC	ET
1881.299	AC	ET
1919.87	AC	
1920.352	AC	
1920.834	AC	
1921.316	AC	ET
1921.798	AC	ET
1944.941	AC	
1945.423	AC	
1945.905	AC	
1946.387	AC	
1996.047	AC	
1996.529	AC	
2016.297	AC	
2016.779	AC	
2017.261	AC	
2017.743	AC	
2018.226	AC	
2042.332	AC	ET
2042.814	AC	ET
2043.297	AC	ET
2043.779	AC	ET
2044.261	AC	
2054.868	AC	
2055.35	AC	
2068.85	AC	
2069.332	AC	
2069.814	AC	
2070.296	AC	
2070.778	AC	
2084.278	AC	
2084.76	AC	
2085.242	AC	
2195.652	AC	
2196.134	AC	
2196.616	AC	
2197.098	AC	
2202.402	AC	
2202.884	AC	
2203.366	AC	
2203.848	AC	
2204.33	AC	
2204.813	AC	
2205.295	AC	

2225.062	AC	
2225.544	AC	
2226.027	AC	
2226.509	AC	
2236.634	AC	ET
2237.116	AC	ET
2237.598	AC	ET
2256.401	AC	
2256.883	AC	
2257.365	AC	
2274.24	AC	
2274.722	AC	
2275.205	AC	
2275.687	AC	
2276.169	AC	
2276.651	AC	
2278.579	AC	
2284.365	AC	ET
2284.847	AC	ET
2285.329	AC	ET
2285.812	AC	ET
2297.383	AC	
2297.865	AC	
2298.347	AC	
2305.097	AC	AC
2305.579	AC	AC
2306.061	AC	AC
2306.543	AC	AC
2307.026	AC	AC
2322.936	AC	
2334.99	AC	
2335.472	AC	
2335.954	AC	
2336.436	AC	
2336.918	AC	
2340.775	AC	AC
2341.257	AC	AC
2341.74	AC	AC
2342.221	AC	AC
2342.704	AC	AC
2343.186	AC	AC
2343.668	AC	AC
2344.15	AC	AC
2344.632	AC	AC
2345.114	AC	AC
2345.596	AC	AC

2360.061	AC	
2360.543	AC	
2361.025	AC	ET
2361.507	AC	ET
2361.989	AC	ET
2368.739	AC	
2369.221	AC	
2369.703	AC	
2370.186	AC	
2370.668	AC	
2371.15	AC	
2371.632	AC	
2372.114	AC	
2372.596	AC	
2373.078	AC	
2373.561	AC	
2374.042	AC	
2374.525	AC	
2375.007	AC	
2375.489	AC	
2375.971	AC	
2376.453	AC	
2376.935	AC	
2377.417	AC	
2377.9	AC	
2378.382	AC	
2394.292	AC	
2394.774	AC	
2395.257	AC	
2395.739	AC	
2427.078	AC	
2462.274	AC	
2462.756	AC	
2463.238	AC	ET
2463.72	AC	ET
2467.095	AC	
2467.577	AC	
2468.059	AC	
2501.809	AC	
2502.291	AC	
2502.773	AC	
2503.255	AC	ET
2510.969	AC	ET
2511.452	AC	ET
2511.934	AC	ET
2512.416	AC	ET

2549.058	AC	AC
2549.541	AC	AC
2550.023	AC	AC
2566.897	AC	AC
2567.38	AC	AC
2567.862	AC	AC
2568.344	AC	AC
2577.022	AC	ET
2577.504	AC	ET
2592.933	AC	ET
2593.415	AC	ET
2593.897	AC	ET
2645.004	AC	AC
2645.486	AC	
2645.968	AC	
2685.985	AC	AC
2686.468	AC	AC
2686.949	AC	AC
2694.182	AC	AC
2694.664	AC	AC
2699.485	AC	AC
2699.967	AC	AC
2700.449	AC	AC
2700.932	AC	AC
2701.414	AC	AC
2701.896	AC	AC
2702.378	AC	AC
2702.86	AC	AC
2711.056	AC	AC
2711.539	AC	AC
2712.021	AC	AC
2712.503	AC	AC
2712.985	AC	AC
2713.467	AC	AC
2728.413	AC	ET
2728.896	AC	ET
2729.378	AC	ET
2729.86	AC	ET
2739.502	AC	ET
2745.288	AC	ET
2745.77	AC	ET
2746.252	AC	ET
2752.038	AC	
2756.377	AC	AC
2756.859	AC	AC
2757.342	AC	AC

2757.824	AC	AC
2758.306	AC	AC
2758.788	AC	AC
2759.27	AC	AC
2959.839	AC	AC
2960.321	AC	AC
2960.803	AC	AC
2961.285	AC	AC
3004.678	AC	ET
3034.57	AC	
3035.052	AC	
3035.535	AC	
3045.659	AC	
3046.142	AC	
3078.927	AC	ET
3079.409	AC	ET
3110.266	AC	ET
3130.998	AC	
3131.48	AC	ET
3163.301	AC	AC
3163.783	AC	AC
3231.282	AC	ET
3231.764	AC	ET
3249.604	AC	
3250.085	AC	
3250.568	AC	
3251.05	AC	
3257.8	AC	AC
3258.282	AC	AC
3258.764	AC	
3327.71	AC	AC
3328.192	AC	AC
3328.674	AC	AC
3329.156	AC	AC
3359.531	AC	
3360.013	AC	
3360.495	AC	AC
3360.977	AC	AC
3400.03	AC	ET
3400.512	AC	ET
3400.994	AC	ET
3401.477	AC	ET
3533.1	AC	
3533.582	AC	ET
3564.439	AC	ET
3564.921	AC	ET

3565.403	AC	ET
3573.117	AC	
3573.599	AC	
3574.082	AC	
3574.564	AC	
3575.046	AC	
3576.974	AC	
3577.457	AC	
3577.939	AC	
3578.421	AC	
3602.528	AC	AC
3603.01	AC	AC
3603.492	AC	AC
3605.903	AC	
3606.385	AC	
3606.867	AC	AC
3607.349	AC	AC
3609.76	AC	AC
3614.581	AC	AC
3615.063	AC	AC
3615.545	AC	AC
3616.028	AC	AC
3617.956	AC	AC
3618.438	AC	AC
3618.92	AC	AC
3619.403	AC	AC
3619.885	AC	AC
3621.331	AC	AC
3621.813	AC	AC
3622.295	AC	AC
3622.777	AC	AC
3623.26	AC	AC
3647.366	AC	ET
3659.42	AC	ET
3659.902	AC	ET
3660.384	AC	ET
3660.866	AC	ET
3661.348	AC	ET
3662.795	AC	ET
3663.277	AC	ET
3663.759	AC	ET
3664.241	AC	ET
3664.723	AC	ET
3667.616	AC	ET
3668.098	AC	ET
3671.473	AC	ET

3673.402	AC	ET
3673.884	AC	ET
3674.366	AC	ET
3674.848	AC	ET
3675.33	AC	ET
3677.741	AC	ET
3678.223	AC	ET
3678.705	AC	ET
3684.009	AC	ET
3689.312	AC	
3689.794	AC	
3690.277	AC	
3694.616	AC	AC
3705.223	AC	AC
3705.705	AC	AC
3706.187	AC	AC
3708.598	AC	AC
3709.08	AC	AC
3709.562	AC	AC
3710.044	AC	AC
3710.526	AC	AC
3711.009	AC	AC
3713.419	AC	AC
3715.83	AC	AC
3716.312	AC	AC
3716.794	AC	AC
3718.24	AC	AC
3718.723	AC	AC
3719.205	AC	
3719.687	AC	
3734.151	AC	AC
3734.633	AC	AC
3746.687	AC	ET
3747.169	AC	ET
3749.58	AC	
3750.062	AC	
3750.544	AC	
3751.026	AC	
3753.919	AC	
3754.401	AC	
3754.883	AC	
3755.365	AC	
3757.776	AC	ET
3758.258	AC	ET
3760.187	AC	
3760.669	AC	

3761.151	AC	
3763.562	AC	
3766.454	AC	
3768.383	AC	ET
3768.865	AC	ET
3769.347	AC	ET
3798.275	AC	ET
3798.758	AC	ET
3805.025	AC	
3806.954	AC	AC
3807.436	AC	
3810.329	AC	
3810.811	AC	
3811.293	AC	
3832.025	AC	
3838.293	AC	
3840.221	AC	
3840.703	AC	ET
3843.114	AC	
3843.596	AC	
3866.739	AC	
3867.221	AC	
3867.703	AC	
3873.007	AC	
3873.489	AC	
3874.935	AC	
3875.417	AC	
3875.899	AC	
3876.382	AC	
3876.864	AC	
3883.614	AC	ET
3884.096	AC	
3884.578	AC	
3885.06	AC	
3885.542	AC	
3886.024	AC	
3895.185	AC	
3895.667	AC	
3905.31	AC	
3910.613	AC	
3911.095	AC	
3911.578	AC	
3912.06	AC	
3916.399	AC	
3916.881	AC	
3917.363	AC	

3917.845	AC	
3919.774	AC	
3920.256	AC	
3920.738	AC	
3921.22	AC	
3921.702	AC	
3927.006	AC	
3927.488	AC	
3927.97	AC	
3928.452	AC	
3928.935	AC	
3931.345	AC	
3931.827	AC	
3935.202	AC	
3935.684	AC	ET
3936.167	AC	ET
3945.327	AC	
3945.809	AC	
3946.292	AC	
3946.773	AC	
3947.256	AC	
3947.738	AC	
3953.523	AC	
3954.488	AC	AC
3954.97	AC	AC
3955.452	AC	AC
3955.934	AC	AC
3956.898	AC	ET
3957.381	AC	ET
3957.863	AC	ET
3958.345	AC	
3958.827	AC	
3959.309	AC	
3959.791	AC	
3960.273	AC	
3962.684	AC	
3963.166	AC	
3963.648	AC	
3964.13	AC	
3964.613	AC	
3965.095	AC	ET
3965.577	AC	ET
3970.88	AC	
3971.363	AC	
3971.845	AC	
3972.327	AC	

3972.809	AC	
3973.291	AC	
3973.773	AC	
3974.255	AC	AC
3974.738	AC	AC
3975.219	AC	AC
3975.702	AC	AC
3976.184	AC	
3976.666	AC	
3977.148	AC	
3978.112	AC	
3978.594	AC	
3979.077	AC	
3979.559	AC	
3980.041	AC	
3980.523	AC	
3981.005	AC	
3981.487	AC	
3981.969	AC	
3982.452	AC	
3982.934	AC	
3983.416	AC	
3983.898	AC	
3984.38	AC	
3984.862	AC	
3985.344	AC	
3985.827	AC	
3986.309	AC	
3986.791	AC	
3987.273	AC	
3987.755	AC	
3988.237	AC	
3988.719	AC	
3989.202	AC	
3989.684	AC	
3990.166	AC	
3990.648	AC	
3991.13	AC	
3991.612	AC	
3992.094	AC	
3992.576	AC	
3993.059	AC	
3993.541	AC	AC
3994.023	AC	AC
3994.505	AC	AC
3994.987	AC	AC

3995.469	AC	AC
3995.951	AC	AC
3996.434	AC	AC
3996.916	AC	AC
3997.398	AC	AC
3997.88	AC	AC
3998.362	AC	AC
3998.844	AC	AC
3999.326	AC	AC
3999.809	AC	AC
4000.291	AC	AC
426.6913	ET	ET
427.1734	ET	ET
427.6555	ET	
428.1377	ET	
428.6198	ET	
429.102	ET	
487.4405	ET	
487.9227	ET	
598.332	ET	ET
598.8141	ET	ET
612.7961	ET	
613.7604	ET	AC
614.2426	ET	AC
614.7247	ET	AC
853.3825	ET	AC
853.8646	ET	AC
854.3468	ET	AC
859.6503	ET	AC
864.9538	ET	AC
865.4359	ET	AC
865.9181	ET	AC
875.5608	ET	AC
876.043	ET	
876.5251	ET	
877.0072	ET	ET
877.4894	ET	ET
877.9715	ET	ET
878.4536	ET	ET
878.9358	ET	ET
879.4179	ET	ET
879.9	ET	ET
883.7571	ET	ET
884.2393	ET	ET
884.7214	ET	ET
885.2036	ET	ET

885.6857	ET	ET
905.9354	ET	AC
906.4176	ET	AC
906.8997	ET	AC
907.3818	ET	AC
907.864	ET	AC
909.7925	ET	AC
910.2747	ET	AC
937.7565	ET	AC
938.2386	ET	AC
938.7208	ET	AC
995.1308	ET	ET
995.6129	ET	ET
996.095	ET	ET
996.5772	ET	ET
997.0593	ET	ET
997.5414	ET	ET
998.0236	ET	ET
998.5057	ET	ET
1007.666	ET	ET
1008.148	ET	ET
1008.631	ET	ET
1009.113	ET	ET
1009.595	ET	ET
1010.077	ET	ET
1010.559	ET	ET
1011.041	ET	ET
1011.523	ET	ET
1012.006	ET	ET
1012.488	ET	ET
1012.97	ET	ET
1013.452	ET	ET
1013.934	ET	ET
1014.416	ET	ET
1014.898	ET	ET
1015.381	ET	ET
1015.863	ET	
1016.345	ET	
1022.13	ET	AC
1022.613	ET	AC
1023.095	ET	AC
1023.577	ET	AC
1024.059	ET	AC
1024.541	ET	AC
1025.023	ET	AC
1025.505	ET	AC

1025.988	ET	AC
1026.47	ET	AC
1026.952	ET	AC
1027.434	ET	AC
1027.916	ET	AC
1028.398	ET	AC
1028.88	ET	AC
1029.363	ET	AC
1029.845	ET	AC
1030.327	ET	AC
1030.809	ET	AC
1031.291	ET	AC
1031.773	ET	AC
1032.255	ET	AC
1032.738	ET	AC
1033.22	ET	AC
1033.702	ET	AC
1034.184	ET	AC
1034.666	ET	AC
1035.148	ET	AC
1035.63	ET	AC
1036.112	ET	AC
1036.594	ET	AC
1037.077	ET	AC
1037.559	ET	AC
1038.041	ET	AC
1039.969	ET	AC
1040.452	ET	AC
1040.934	ET	AC
1041.416	ET	AC
1041.898	ET	AC
1042.38	ET	AC
1042.862	ET	AC
1043.344	ET	AC
1043.827	ET	AC
1044.309	ET	
1044.791	ET	ET
1046.719	ET	ET
1047.202	ET	ET
1047.684	ET	ET
1048.166	ET	ET
1048.648	ET	ET
1049.13	ET	ET
1049.612	ET	ET
1050.094	ET	ET
1050.577	ET	ET

1051.059	ET	ET
1051.541	ET	ET
1052.023	ET	ET
1052.505	ET	ET
1052.987	ET	ET
1053.469	ET	ET
1053.951	ET	ET
1054.434	ET	ET
1054.916	ET	ET
1055.398	ET	ET
1055.88	ET	ET
1056.362	ET	ET
1056.844	ET	ET
1057.326	ET	ET
1057.809	ET	ET
1058.291	ET	ET
1058.773	ET	ET
1065.523	ET	AC
1066.005	ET	AC
1066.487	ET	AC
1066.969	ET	AC
1067.451	ET	AC
1067.933	ET	AC
1068.416	ET	AC
1068.898	ET	AC
1069.38	ET	AC
1069.862	ET	AC
1070.344	ET	AC
1070.826	ET	AC
1071.308	ET	AC
1071.791	ET	AC
1072.273	ET	AC
1072.755	ET	AC
1073.237	ET	AC
1073.719	ET	AC
1074.201	ET	AC
1076.13	ET	AC
1076.612	ET	AC
1077.094	ET	AC
1077.576	ET	AC
1078.058	ET	AC
1078.541	ET	AC
1079.023	ET	AC
1079.505	ET	AC
1079.987	ET	AC
1080.469	ET	AC

1080.951	ET	AC
1081.433	ET	AC
1081.916	ET	AC
1082.398	ET	AC
1082.88	ET	AC
1083.362	ET	AC
1083.844	ET	AC
1084.326	ET	AC
1084.808	ET	AC
1085.29	ET	AC
1085.772	ET	ET
1086.737	ET	ET
1087.219	ET	ET
1087.701	ET	ET
1088.183	ET	ET
1088.665	ET	ET
1089.147	ET	ET
1089.63	ET	ET
1090.112	ET	ET
1090.594	ET	ET
1091.076	ET	ET
1091.558	ET	ET
1092.04	ET	ET
1092.522	ET	ET
1093.005	ET	ET
1093.487	ET	ET
1093.969	ET	ET
1094.451	ET	ET
1094.933	ET	ET
1095.415	ET	ET
1095.897	ET	ET
1096.38	ET	ET
1096.862	ET	ET
1097.344	ET	ET
1097.826	ET	ET
1098.308	ET	ET
1098.79	ET	ET
1099.272	ET	ET
1099.755	ET	ET
1100.237	ET	ET
1100.719	ET	ET
1101.201	ET	ET
1101.683	ET	ET
1102.165	ET	ET
1108.433	ET	ET
1108.915	ET	

1120.004	ET	AC
1120.486	ET	AC
1120.969	ET	AC
1121.451	ET	AC
1121.933	ET	AC
1122.415	ET	AC
1122.897	ET	AC
1123.379	ET	AC
1123.861	ET	AC
1161.95	ET	ET
1206.307	ET	AC
1206.789	ET	AC
1207.271	ET	AC
1207.753	ET	AC
1208.235	ET	AC
1208.718	ET	AC
1231.378	ET	ET
1244.878	ET	ET
1245.36	ET	ET
1245.842	ET	ET
1246.324	ET	ET
1246.806	ET	ET
1247.288	ET	ET
1258.378	ET	
1258.86	ET	
1259.342	ET	
1259.824	ET	
1295.02	ET	
1295.502	ET	
1295.984	ET	
1296.466	ET	
1296.948	ET	AC
1297.431	ET	AC
1297.913	ET	AC
1298.395	ET	AC
1298.877	ET	AC
1299.359	ET	AC
1299.841	ET	AC
1300.323	ET	AC
1300.806	ET	AC
1301.288	ET	AC
1301.77	ET	AC
1302.252	ET	AC
1302.734	ET	AC
1303.216	ET	AC
1303.698	ET	AC

1304.181	ET	AC
1304.663	ET	AC
1308.038	ET	AC
1308.52	ET	AC
1310.448	ET	AC
1310.93	ET	AC
1311.413	ET	AC
1311.895	ET	AC
1312.377	ET	AC
1312.859	ET	AC
1313.341	ET	AC
1313.823	ET	AC
1314.305	ET	AC
1314.788	ET	AC
1315.27	ET	AC
1315.752	ET	
1316.234	ET	
1319.127	ET	ET
1319.609	ET	ET
1320.091	ET	ET
1320.573	ET	ET
1321.055	ET	ET
1321.538	ET	ET
1322.02	ET	ET
1322.502	ET	ET
1322.984	ET	ET
1323.466	ET	ET
1323.948	ET	ET
1324.43	ET	ET
1324.912	ET	ET
1325.395	ET	
1325.877	ET	
1326.359	ET	
1326.841	ET	
1327.323	ET	AC
1327.805	ET	AC
1328.287	ET	AC
1328.77	ET	AC
1329.252	ET	AC
1336.002	ET	AC
1336.484	ET	AC
1360.108	ET	ET
1360.591	ET	ET
1361.073	ET	ET
1361.555	ET	ET
1362.037	ET	ET

1362.519	ET	ET
1363.001	ET	ET
1363.483	ET	ET
1400.126	ET	AC
1400.608	ET	AC
1401.09	ET	AC
1401.572	ET	AC
1402.054	ET	AC
1402.537	ET	AC
1403.019	ET	AC
1403.501	ET	AC
1403.983	ET	AC
1404.465	ET	AC
1404.947	ET	AC
1405.429	ET	AC
1405.911	ET	AC
1406.394	ET	AC
1406.876	ET	AC
1407.358	ET	AC
1407.84	ET	AC
1408.322	ET	AC
1408.804	ET	AC
1409.286	ET	AC
1409.769	ET	AC
1410.251	ET	AC
1410.733	ET	AC
1411.215	ET	AC
1411.697	ET	AC
1412.179	ET	AC
1412.661	ET	AC
1413.144	ET	AC
1413.626	ET	AC
1424.233	ET	ET
1424.715	ET	ET
1425.197	ET	ET
1427.125	ET	ET
1427.608	ET	ET
1428.09	ET	ET
1428.572	ET	ET
1429.054	ET	ET
1429.536	ET	ET
1430.018	ET	ET
1430.5	ET	ET
1430.983	ET	ET
1431.465	ET	ET
1431.947	ET	ET

1432.429	ET	ET
1432.911	ET	ET
1433.393	ET	ET
1448.34	ET	AC
1448.822	ET	AC
1449.304	ET	AC
1449.786	ET	AC
1450.268	ET	AC
1450.75	ET	AC
1451.232	ET	AC
1464.732	ET	ET
1465.214	ET	ET
1541.874	ET	AC
1542.356	ET	AC
1542.838	ET	AC
1586.231	ET	ET
1586.713	ET	ET
1587.195	ET	ET
1626.248	ET	AC
1626.73	ET	AC
1627.212	ET	AC
1627.695	ET	AC
1628.177	ET	AC
1628.659	ET	AC
1629.141	ET	AC
1629.623	ET	AC
1630.105	ET	AC
1630.587	ET	AC
1631.07	ET	AC
1631.552	ET	AC
1646.98	ET	ET
1647.462	ET	ET
1647.944	ET	ET
1648.427	ET	ET
1648.908	ET	
1649.391	ET	
1656.141	ET	AC
1656.623	ET	AC
1659.516	ET	AC
1659.998	ET	AC
1660.48	ET	AC
1660.962	ET	AC
1661.444	ET	AC
1661.926	ET	AC
1662.408	ET	AC
1662.891	ET	AC

1663.373	ET	AC
1663.855	ET	AC
1664.337	ET	AC
1664.819	ET	AC
1665.301	ET	AC
1665.783	ET	AC
1666.265	ET	AC
1666.748	ET	AC
1667.23	ET	AC
1667.712	ET	AC
1671.087	ET	
1671.569	ET	ET
1672.051	ET	ET
1672.533	ET	ET
1673.015	ET	ET
1679.283	ET	ET
1679.765	ET	ET
1680.248	ET	ET
1680.729	ET	ET
1681.212	ET	ET
1681.694	ET	ET
1682.176	ET	ET
1682.658	ET	ET
1683.14	ET	ET
1683.622	ET	ET
1684.104	ET	ET
1684.587	ET	ET
1685.069	ET	ET
1685.551	ET	ET
1686.033	ET	ET
1686.515	ET	ET
1686.997	ET	ET
1687.479	ET	ET
1687.962	ET	ET
1688.444	ET	ET
1688.926	ET	ET
1689.408	ET	ET
1689.89	ET	ET
1690.372	ET	ET
1690.854	ET	ET
1693.265	ET	ET
1693.747	ET	ET
1694.229	ET	ET
1694.712	ET	ET
1695.194	ET	ET
1695.676	ET	ET

1698.086	ET	ET
1698.569	ET	ET
1699.051	ET	ET
1699.533	ET	ET
1700.015	ET	ET
1700.497	ET	ET
1700.979	ET	ET
1708.211	ET	AC
1708.694	ET	AC
1709.176	ET	AC
1709.658	ET	AC
1710.14	ET	AC
1710.622	ET	AC
1711.104	ET	AC
1711.586	ET	AC
1712.069	ET	AC
1712.551	ET	AC
1713.033	ET	AC
1713.515	ET	AC
1713.997	ET	AC
1714.479	ET	AC
1714.961	ET	AC
1718.818	ET	AC
1719.301	ET	AC
1719.783	ET	AC
1720.265	ET	AC
1720.747	ET	AC
1721.229	ET	AC
1721.711	ET	AC
1722.193	ET	AC
1722.676	ET	AC
1723.158	ET	AC
1723.64	ET	AC
1726.533	ET	AC
1727.015	ET	AC
1727.497	ET	AC
1727.979	ET	AC
1728.461	ET	AC
1728.943	ET	AC
1729.426	ET	AC
1743.407	ET	ET
1743.89	ET	ET
1744.372	ET	ET
1744.854	ET	ET
1745.336	ET	ET
1757.389	ET	ET

1757.872	ET	ET
1758.354	ET	ET
1758.836	ET	ET
1759.318	ET	ET
1765.104	ET	ET
1765.586	ET	ET
1766.068	ET	ET
1766.55	ET	
1767.032	ET	
1767.514	ET	
1797.889	ET	ET
1798.371	ET	ET
1798.853	ET	ET
1851.406	ET	AC
1851.888	ET	AC
1935.78	ET	
1970.494	ET	
1970.976	ET	
2040.886	ET	
2134.42	ET	
2134.903	ET	
2261.222	ET	
2403.453	ET	ET
2403.935	ET	ET
2429.488	ET	
2429.97	ET	
2430.453	ET	AC
2430.935	ET	AC
2598.236	ET	ET
2598.719	ET	ET
2599.201	ET	ET
2599.683	ET	ET
2607.397	ET	ET
2607.879	ET	ET
2608.361	ET	ET
2608.843	ET	ET
2609.325	ET	ET
2609.808	ET	ET
2610.29	ET	ET
2620.897	ET	AC
2621.379	ET	AC
2621.861	ET	AC
2669.593	ET	ET
2670.075	ET	ET
2670.557	ET	ET
2671.039	ET	ET

2671.521	ET	ET
2672.485	ET	ET
2672.968	ET	ET
2707.681	ET	AC
2708.164	ET	AC
2708.646	ET	AC
2709.61	ET	AC
2710.092	ET	AC
2746.735	ET	ET
2747.217	ET	ET
2747.699	ET	ET
2748.181	ET	ET
2795.913	ET	AC
2945.375	ET	ET
2965.143	ET	AC
2965.625	ET	AC
2966.107	ET	AC
2966.589	ET	AC
2967.071	ET	AC
2993.106	ET	ET
2993.589	ET	ET
2994.071	ET	ET
2994.553	ET	ET
2995.035	ET	ET
3001.303	ET	ET
3001.785	ET	ET
3002.267	ET	ET
3002.749	ET	ET
3003.231	ET	ET
3003.714	ET	ET
3004.196	ET	ET
3004.678	ET	ET
3019.624	ET	AC
3020.106	ET	AC
3025.892	ET	AC
3026.374	ET	AC
3026.856	ET	AC
3070.249	ET	ET
3070.731	ET	ET
3071.213	ET	ET
3111.23	ET	
3111.712	ET	
3112.194	ET	
3112.677	ET	
3155.587	ET	AC
3156.069	ET	AC

3156.551	ET	AC
3168.604	ET	
3188.372	ET	AC
3188.854	ET	AC
3209.104	ET	AC
3209.586	ET	AC
3210.068	ET	
3210.55	ET	
3211.032	ET	
3236.104	ET	ET
3239.479	ET	ET
3239.961	ET	ET
3240.443	ET	ET
3240.925	ET	ET
3241.407	ET	ET
3273.71	ET	ET
3274.192	ET	ET
3274.675	ET	ET
3300.71	ET	
3301.192	ET	
3332.049	ET	AC
3340.245	ET	AC
3340.727	ET	AC
3341.209	ET	AC
3341.692	ET	AC
3342.174	ET	AC
3342.656	ET	AC
3343.138	ET	AC
3343.62	ET	AC
3373.031	ET	AC
3394.245	ET	ET
3394.727	ET	ET
3395.209	ET	ET
3781.401	ET	ET
3781.883	ET	ET
3806.472	ET	AC

Fuente: Propia autoría