UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE BIOTECNOLOGÍA

COMPARACIÓN DE TÉCNICAS DE PROCESAMIENTO DE DATOS OBTENIDOS A TRAVÉS DE ESPECTROS EN EL FTIR PARA LA CUANTIFICACIÓN SIMULTÁNEA DE METANOL Y ETANOL PRESENTES EN BEBIDAS ALCOHÓLICAS ARTESANALES PRODUCIDAS EN EL VALLE DE YUNGUILLA

Trabajo de titulación previo para la obtención del

título de Ingeniero Biotecnólogo

AUTOR: RICARDO ESTEBAN JARA MOSCOSO

TUTOR: ING. PABLO WILSON ARÉVALO MOSCOSO, PhD.

Cuenca - Ecuador

2025

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Ricardo Esteban Jara Moscoso con documento de identificación N° 0107895369 manifiesto que:

Soy el autor 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, 24 de septiembre del 2025

Atentamente,

Ricardo Esteban Jara Moscoso

0107895369

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

Yo, Ricardo Esteban Jara Moscoso con documento de identificación N° 0107895369, 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 autor del Trabajo experimental: "Comparación de técnicas de procesamiento de datos obtenidos a través de espectros en el FTIR para la cuantificación simultánea de metanol y etanol presentes en bebidas alcohólicas artesanales producidas en el valle de Yunguilla", el cual ha sido desarrollado para optar por el título de: Ingeniero Biotecnólogo, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

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

Cuenca, 24 de septiembre del 2025

Atentamente,

Ricardo Esteban Jara Moscoso

0107895369

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Pablo Wilson Arévalo Moscoso con documento de identificación N° 0102156957, docente de la Universidad Politécnica Salesiana, declaró que bajo mi tutoría fue desarrollado el trabajo de titulación: COMPARACIÓN DE TÉCNICAS DE PROCESAMIENTO DE DATOS OBTENIDOS A TRAVÉS DE ESPECTROS EN EL FTIR PARA LA CUANTIFICACIÓN SIMULTÁNEA DE METANOL Y ETANOL PRESENTES EN BEBIDAS ALCOHÓLICAS ARTESANALES PRODUCIDAS EN EL VALLE DE YUNGUILLA, realizado por Ricardo Esteban Jara Moscoso con documento de identificación N° 0107895369 , 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, 24 de septiembre del 2025

Atentamente,

Ing. Pablo Wilson Arévalo Moscoso, PhD.

0102156957

ÍNDICE

DEDICATORIA

El presente trabajo está dedicado primordialmente a Dios que siempre me ayuda a despertar cada día, para dar lo mejor de mí.

A mis padres Marcelo e Irma, creadores de mi vida y que nunca dejaron de creer en mí a pesar de todos mis errores.

A mis hermanos, Paulo y Michelle, los cuales fueron centinelas en mi vida, brindándome guía, apoyo, cariño, respeto, siempre han tenido toda mi admiración.

Para Rupert, Kae y Felix quienes me enseñaron como se siente realmente el amor incondicional que solo puede venir por parte un animal.

Para el alcoholismo de mi padre motivó esta investigación.

A mis queridos amigos, que siempre tuvieron el carisma para sacarme risas.

A todos mis compinches de Biotecnológica que pasaron a formar parte de mi familia, agradezco a dios todas las horas de clases, prácticas, paseos y salidas que vivimos.

A todos mis maestros que ayudaron y contribuyeron durante mi formación académica.

A todas las personas a las que pude conocer durante la carrera, que sumaron en mi vida, a todas mujeres que amé durante la carrera y a mis tres más grandes amores María, pucho y al diablo embotellado.

AGRADECIMIENTOS

Agradezco profundamente a dios por permitirme vivir esta hermosa etapa, por ayudarme desde su guía a concluir este trabajo de titulación, además de ser un gran paso para mi carrera profesional, la biotecnología se convirtió una parte importante de mi vida.

Además, debo agradecer profundamente al mi amigo Edmond Geraud más que un profesor o un mentor se convirtió en un gran amigo personal que siempre es una persona sincera, amena, trabajadora y valiosa para toda la comunidad dentro de la UPS, y con su guía y paciencia se logró finalizar el presente trabajo.

Al Doctor Pablo Wilson Arévalo Moscoso le agradezco su infinita paciencia, guía y tutela que con la ayuda de su visión y su experiencia ayudaron a formar la investigación del presente trabajo

Así mismo agradezco a la ingeniera Myriam Mancheno que fue una parte importante dentro de mi formación académica y por la frase que marco mucho en mi vida universitaria: *"Usen su ingenio por eso son ingenieros "*.

Además, un gran agradecimiento a al equipo del departamento de los laboratorios de ciencias de la vida que siempre ayudaron y colaboraron en todas las prácticas y proyectos elaborados.

También agradezco a dios por poner en mi camino a todos los compañeros que después se convirtieron en amigos muy queridos para mí (J.R, R.E, F.L,J.D, R.I).

Y por último agradezco profundamente a mis padres, mis abuelos, mis primos, y mis queridos hermanos, los amo desde el fondo del corazón.

y como dijo Gustavo Cerati

¡ ¡ ¡Gracias totales !!!

RESUMEN

En el presente trabajo se aborda la investigación de tipo explicativa en la que se busca responde a la integrante de sobre qué procesamiento espectral contribuye al desarrollo de dos modelos con capacidad de intuir la cuantificación de etanol y metanol por FTIR, si es mejor la predicción a partir del espectro completo o solo en los picos de interés con el objetivo que se prediga las concentraciones de estos en una muestra de alcohol artesanal elaborado en la provincia de yunguilla. Con el fin de determinar concentraciones de metanol que pueden ser perjudiciales para la salud. Se hizo un estudio experimental en el cual se dividió en dos partes principales, en la primera consistió en la elaboración de estándares de concentración y medición de su onda espectral, para ello se emplearon técnicas como es la espectroscopia en FTIR.

La segunda parte fue el posterior en análisis informático con software especializado, dentro del estudio se llegó a la que la mejor técnica para la elaboración de modelos multilíneas se hace a través de la metodología sPLS que fue formado a partir de estándares en los que se utilizó el espectro completo, y que al momento de evaluar se obtuvo un MSE de 0.360 lo que dictamina una gran precisión en el modelo pero solo va a ser efectivó dentro del rango establecido por el diseño de muestras en caso de que la muestra tenga mayor grado el modelo tiene a extrapolarse restando eficiencia.

ABSTRACT

This work addresses explanatory research that seeks to answer the question of which spectral processing contributes to the development of two models capable of intuiting the quantification of ethanol and methanol by FTIR, whether prediction is better based on the complete spectrum or only on the peaks of interest, with the aim of predicting

the concentrations of these in a sample of artisanal alcohol produced in the province of Yunguilla. The aim is to determine methanol concentrations that may be harmful to health. An experimental study was conducted, divided into two main parts. The first part consisted of developing concentration standards and measuring their spectral wave, using techniques such as FTIR spectroscopy.

The second part was the subsequent computer analysis with specialized software. The study concluded that the best technique for developing multiline models is through the sPLS methodology, which was formed from standards in which the complete spectrum was used. and at the time of evaluation, an MSE of 0.36 of was obtained, which indicates a high degree of accuracy in the model, but it will only be effective within the range established by the sample design. If the sample has a higher degree, the model tends to extrapolate, reducing efficiency.

CAPÍTULO 1

1.1 Planteamiento del problema

En los interiores de la provincia del Azuay, se ubica el cantón de Santa Isabel a aproximadamente 70 km de la ciudad de Cuenca. En él se ubica el valle de Yunguilla, este es característico porque se encuentra en un microclima privilegiado, tiene clima cálido húmedo casi todo el año entre temperaturas que bordean los 20 a 25 grados centígrados, gracias a esta característica, muchos habitantes de la zona viven a través de la agricultura como su actividad económica principal. Una de las especies más cultivadas en la zona es el cultivo de caña de azúcar (*Saccharum Officinarum)* como se puede apreciar en la figura 1. Así mismo muchos otros habitantes también se dedican a la transformación de esta materia prima en productos con valor agregado, entre ellos figuran la producción de Panela, y alcohol principalmente.

Figura 1: Parcela cultivada con caña de azúcar dentro del valle de yunguilla

Imagen elaborada por el Autor.

El problema nace durante en el proceso de transformación de la caña de azúcar a productos de valor agregado ya que los métodos e instrumentos con los que se lo realiza son artesanales sin medidas de control ni asepsia, además que no cumplen con ningunas normas de salubridad requeridas para esos productos. En procesos más sencillos como la elaboración de Panela existe un riesgo al consumirla, pero es menor que al comparar el consumo de alcohol de origen artesanal que sí está mal destilado este puede llegar a ser tóxico para los humanos causando desde cegara a ésta en casos más graves coma e incluso la muerte.

Figura 2: Trapiche para la extracción del jugo de caña para la posterior, fermentación y destilación.

Imagen elaborada por el Autor.

Para que un Alcohol artesanal tenga esta capacidad tóxica para un ser humano se da principalmente por a la hora de destilar con ayuda un trapiche el cual por presión mecánica se extrae el jugo de caña y pasa a un proceso fermentativo que generalmente suele durar entre 3 a 5 días, al cumplir este tiempo se lo destila con ayuda de un alambique el cual se calienta el fermento y mediante diferencia de puntos de ebullición se destila el etanol a 73 grados centígrados, pasa a través de un serpentín que enfría y condensa el alcohol para pasar a su posterior filtración y reposo para su posterior consumo.

Dentro del proceso de extraer el jugo de caña a partir del bagazo se liberan diferentes pectinas las cuales no se forma de la nada, sino que dependiendo de las características de la especie y la condición del cultivo existe mayores o menores concentraciones. Estas al fermentar se transforman en metanol, por ende, entre más concentración de metanol mayor la toxicidad este es recuperado a los 65 Cº por lo que es lo primero que se destilara. Por otro lado, un punto primordial para evitar la mezcla de metanol en el alcohol final es necesario personal capacitado que sepa bien a las temperaturas en las que ya es seguro recuperar el destilado, por lo que no es raro que durante un proceso de destilación fraccionada el alcohol obtenido suele estar contaminado (Juan V. Ashurst et al., 2025).

Dentro del cuerpo humano el problema no es el compuesto, sino la forma en que el cuerpo lo trata de deshacer del mismo, el metanol al ingerirlo es absorbido e ingresa al torrente sanguíneo hasta el hígado en donde va a ser metabolizado. Para poder eliminarlo el hígado lo transforma en formaldehido con ayuda de una enzima alcohol deshidrogenasa y a su vez este nuevo compuesto se puede transformar en ácido fórmico, gracias a la enzima aldehído deshidrogenasa. Estos intermediarios metabólicos son altamente tóxicos

para el cuerpo y generan problemas a nivel celular en el proceso de la respiración celular. A nivel sistémico en el cuerpo puede llevar a causar una acidosis metabólica. (Villanueva Anadón et al., 2002).

La principal sintomatología que consumidores puede generar son fácilmente confundibles con una intoxicación etílica común dado que comparten muchos síntomas: dolor de cabeza, dolor abdominal, sensación de embriaguez, pero después de haber trasformado el metanol se empiezan a ver las diferencias, dado que si hubo una intoxicación media por metanol los síntomas se presentan como: sensibilidad visual, fotofobia, ceguera parcial o total y en casos de intoxicación extrema se puede producir una acidosis metabólica, coma, convulsiones e inclusive la muerte (Juan V. Ashurst et al., 2025).

En ecuador este es un problema que se ha repetido muchas veces de forma persistente en distintos sectores del país, acorde a el ministerio de salud pública un caso que ocurrió el 19 de noviembre de 2022, En el cual se atendieron a 136 personas con sospecha de intoxicación por el consumo de alcohol metílico (Ministerio de Salud Pública, 2020).De este grupo, 17 de ellos fallecieron meses posteriores, otro caso de intoxicación masiva se presentó el 3 de diciembre del mismo año, en el que se han atendido 178 personas con sospecha de intoxicación por alcohol metílico, con localización en Quito (159) y Latacunga (19).(Ministerio de Salud Pública, n.d.).

1.2 Pregunta de investigación

¿Qué técnica de procesamiento mejora la precisión para la cuantificación simultánea de etanol y metanol a través de FTIR en bebidas alcohólicas de origen artesanal?

1.3 Justificación de la investigación

El consumo de bebidas alcohólicas en sí ya es perjudicial para la salud humana, ahora aumentándole otros problemas como lo es un origen turbulento de la bebida alcohólica eso puede inducir a un aumento en el número de casos por posibles intoxicaciones por metanol causado por la irregularidad de estos productos, por eso la calidad y seguridad de los productos es vital para los consumidores, así protegiendo a la población y evitando posibles problemas de salud pública.

Para el análisis de alcoholes según las normas INEN los niveles de alcohol del permitido dentro para el aguardiente de caña está entre los 10 mg por cada 100 ml es decir inferior al 0,1%, para su cuantificación el método preferente o ideal acorde a las normas INEN para la identificación es la cromatografía de gases, aunque su principal inconveniente radica en los altos costos y el requerimiento de equipo especial. de además existe otro método el cual contribuye ayudar para el análisis es la titulación, pero el principal inconveniente de esta técnica son que requieren, altos costos, personal especializado, requieren de tiempo para obtener los resultados, por lo que se puede optar por otra opción como es la espectroscopía infrarroja con transformada de Fourier (FTIR)

en la cual podemos ver un espectro en el cual cada pico corresponde a la absorción de energía por parte de un enlace químico (INEN, 2013).

Con la ayuda de esta herramienta y con la formación de modelos matemáticos a partir del análisis de los espectros producidos por el equipo. Permite establecer modelos que solo con la ayuda de los espectros que es un análisis más rápido, fácil y menos costos podemos obtener resultados tentativos rápidos, confiables y más eficientes sobre la cuantificación de las sustancias que deterioran la salud de los consumidores, así reduciendo el número de caso por intoxicaciones. También permite garantizar que la norma INEN se cumpla. Además, mediante el modelamiento aumenta la posibilidad de poder extrapolar la investigación hacia otras industrias con otras variables.

1.4 Limitaciones

Las principales limitaciones de la investigación son la falta de estandarización del alcohol dado que durante cada proceso de destilación existe cambios que afectan las concentraciones limitado a muestras representativas, además dado que cada persona que destile va a tener su proceso definido por lo que las muestras de bebidas alcohólicas siempre varían.

Otra limitante importante es causada por las limitantes espectrales se puede presentar en el equipo ya será por la muestra presentada o la calibración del equipo y estas van a ser catalogadas como posibles interferidas espectrales que va a afecta a los datos

espectrales que van a formar parte del moldeo, último factor a considerar es el tiempo limitado lo que imposibilita un análisis más exhaustivo de las muestras.

1.5 Objetivos

1.5.1 Objetivo General :

Determinar la técnica de análisis de datos más adecuada para lograr la mejor predicción en la cuantificación simultánea de metanol y etanol, a partir de los espectros obtenidos mediante FTIR en bebidas alcohólicas artesanales producidas en el valle de Yunguilla.

1.5.2 Objetivo específico:

- Determinar la interrelación entre las concentraciones de etanol - metanol y sus espectros.

- Evaluar modelos predictivos utilizando datos obtenidos de los espectros para realizar predicciones simultáneas de las concentraciones de etanol y metanol.

- Evaluar los datos obtenidos del modelo en comparación con los valores reales presentes en la muestra

1.7 Hipótesis

El uso de técnicas de procesamiento de los espectros generados a partir del FTIR permite una mejora en el modelo predictivo en la cuantificación del etanol y metanol en bebidas de origen artesanal.

Capítulo 2

2.1 Marco teórico

2.1.1 Antecedentes de la investigación

Con el desarrollo de diferentes técnicas se ha facilitado muchos procesos, con la espectroscopia FTIR esta se convirtió en una técnica que facilita muchos aspectos, reducen costo y tiempo además de su versatilidad en muchos campos de aplicación, de manera general esta técnica se basa en poder identificar tentativamente los grupos funcionales presentes en la muestra estos son identificados por la un haz de luz infrarroja que índice sobre la muestra y con un sensor registra cambios de la vibraciones de sus enlaces y genera un espectro.(Berthomieu & Hienerwadel, 2009a). Con esta ventaja a través de los años se volvió una herramienta clave en múltiples estudios que emplean técnicas de espectroscopia infrarroja.

En el caso del estudio igual enfocado en la evaluación alcoholes de desarrollado por se muestra un estudio en el cual mediante espectroscopia de infrarrojo y técnicas discriminantes como DA, PCA, SVM y SIMCA, realizaron los análisis multivariantes para la discriminación de 110 vinos en función acorde a la variedad de uva y su composición. Dentro del estudio se emplearon un rango espectral (4000-700cm-1), los resultados de más adecuado fueron obtenidos con el tratamiento con SVM con una clasificación exitosa superior al 72.2% en los sets de entrenamiento y un 44% para el set de validación.(Manuel Vázquez Vázquez, n.d.).

Otra aplicación de la espectroscopia se da en el estudio de(Yaman, 2020), en el cual se empleó espectroscopia de Raman e infrarroja. el objetivo principal de estudio de este fue detectar muestras de leche vaca que se la hacían pasar como leche de cabra, en el estudio se usaron una proteína (B-caroteno) que solo estaba presente dentro de las muestras de origen vacuno a través de un modelo de PLS y este alcanzó a una correlación de validación del 0.96.

Asimismo, en el estudio por parte de (Da Silva et al., 2020) en el cual se centraron en desarrollar un método analítico automatizado que tenga la capacidad para caracterizar micro plásticos de un tamaño inferior a 100 micras para los plásticos más usados dentro de la industria, en el estudio se emplearon imágenes hiperespectrales por micro transformada de Fourier combinado con herramientas de Deep learning como PLS-DA y SIMCA. Siendo el mejor modelo de PLS-DA con tuvo la capacidad de generar una clasificación con un margen de error del 1% y con una tasa de clasificación de éxito del 95 %. La muestra usada dentro del estudio fue a partir de sedimentos de diferentes mezclas de plástico.

Dentro de otros campos de investigación también se lo ha usado de técnica como es dentro de la industria alimentaria como es el caso de un estudio en Indonesia en el motivo de la investigación fue la elaboración de un modelo PLS discriminatorio que tenga la capacidad de identificar carne de rata dentro de muestras de carne molida adulteradas. Para la elaboración del modelo emplearon los espectros formados a partir del tipo de grasa entre una muestra de rata y otra de cerdo. Y se obtuvo un modelo que a partir de una

muestra extraña se determine la presencia y cantidad de carne de rata.(Rahmania et al., 2015).

En otro estudio de igual manera relacionado al área de adulteración de alimentos se buscó discernir entre muestras de miel y miel agregada de azúcares externos. con el fin de distinguir entre una muestra verdadera y una adúltera se elaboró mediante ATR-FTIR y a través de procesos quimiométricos, lo que indico como es un método totalmente confiable para fines de control de calidad.(Sahlan et al., 2019)

En ámbitos relacionados a la biotecnología ambiental tiene aplicaciones importantes para la salud de los suelos como lo indica el autor del estudio(Devianti et al., 2019). en el cual se busca identificar metales pesados (Zn y Pb) en suelos agrícolas con el fin de evitar la contaminación en los cultivos, para ello se elaboró un modelo PCA que obvio un margen de error de 3% para la distensión entre presencia o ausencia y para la predicción simultánea consiguieron una correlación de 0.98 lo que indica una alta capacidad predictiva.

Dentro de una rama completamente diferente, en el estudio elaborado por (Smok-Kalwat et al., 2024) . detalla la elaboración de un análisis (PCA), en el cual mediante un código automático de machine learning, máquinas de soporte vectorial (SSVM) se desarrolló un modelo predictivo con el propósito de mediante espectros del FTIR con un rango número de onda de 800 -1800cm-1 permite la detección de cánceres, mediante muestras liquidas de tejidos, siendo un uso potencial de la técnica dentro de la industria médica.

Después de esta investigación se puede afirmar la versatilidad de esta técnica y como tiene tantas aplicaciones en diferentes campos y como es escalable en distintas técnicas.

2.2 Bases teóricas

2.2.1 Espectroscopia infrarroja

La espectroscopia infrarroja (IR) es una técnica que interrelaciona las áreas de física y química en el cual se busca ver el comportamiento de las moléculas químicas cuando son sometidas a un haz de radiación, al momento de que las moléculas fueron irradiadas, dependiendo de su naturaleza y tipo de compuesto y enlace formado se van a generar diversas excitaciones vibracionales únicas para cada tipo grupo funcional y enlace. (FTIR | FTIR Spectroscopy Academy | Thermo Fisher Scientific - IE, n.d.).

La excitación vibracional de cada tipo de compuesto ocurre debió a que pasa un fotón a través de la molécula y si tiene exactamente la cantidad de energía justa y necesaria el enlace absorbe y provoca una vibración que luego es detectada y ese pulso se lo traduce a un espectro para ser analizado.(Athavale et al., 2020)

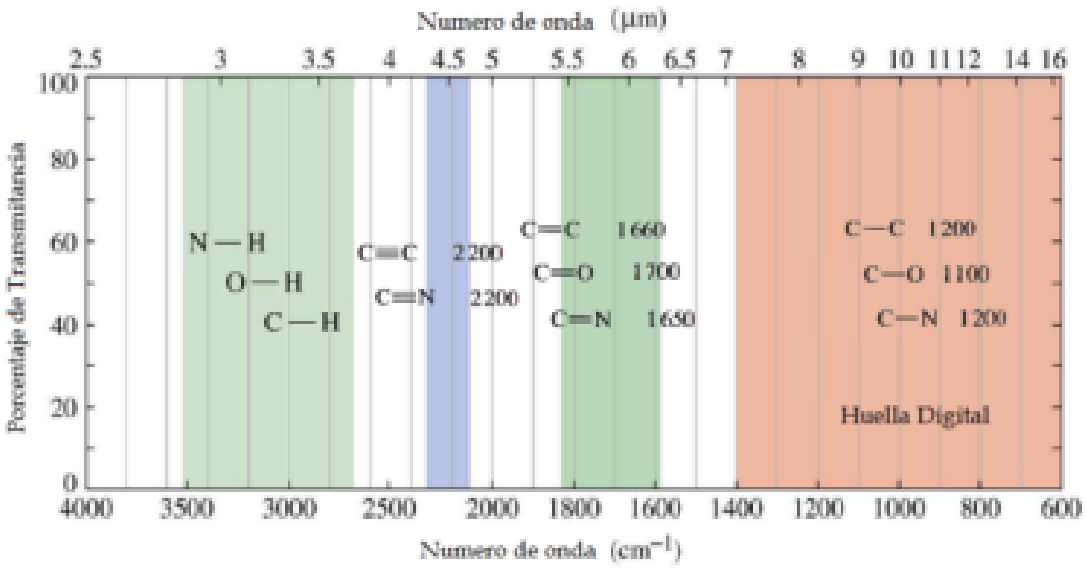
Los espectros generados son presentados en términos de números de onda, que se relaciona con la frecuencia de la radiación que absorbió mediante la siguiente ecuación:

$$\tilde{V} = V/c$$

V es la frecuencia y C es la constante de la velocidad de la luz (Fleming Patrick, n.d.)

2.2.2 Región diagnóstica y región de las huellas digitales

En un espectro de (IR) se da la información en picos ordenados, se divide en dos partes el espectro, una de región de diagnóstico en el cual indica información clara y precisa que dependiendo de que altura de la número de onda en la que se forma el pico puede ser un compuesto u otro, en contraparte la región de huella genética corresponde a la número de onda de entre 1500 y 500, en esta zona existen muchas señales que producen un patrón complejo lo que complica el análisis, sin embargo es especialmente útil para comparar o confirmar espectros (Document, n.d.).



Fuente: *Capitulo 2 Espectroscopia del infrarrojo 2.1 Región Del Infrarrojo.*

Figura 3: *En la siguiente imagen se ve cómo se divide las regiones del espectro y dependiendo de la número de onda se aprecia uno u otro enlace.*

2.2.3 Señal residual Background

La señal residual del background o también llamado ruido esta es una función consiste en la eliminación del a señal que queda antes del análisis de una muestra, esto con el fin que la señal de la muestra no posea interferencias entre el fondo en donde está el equipo y los picos de interés.(CAPÍTULO II Espectroscopia Del Infrarrojo 2.1 Región Del Infrarrojo, n.d.).

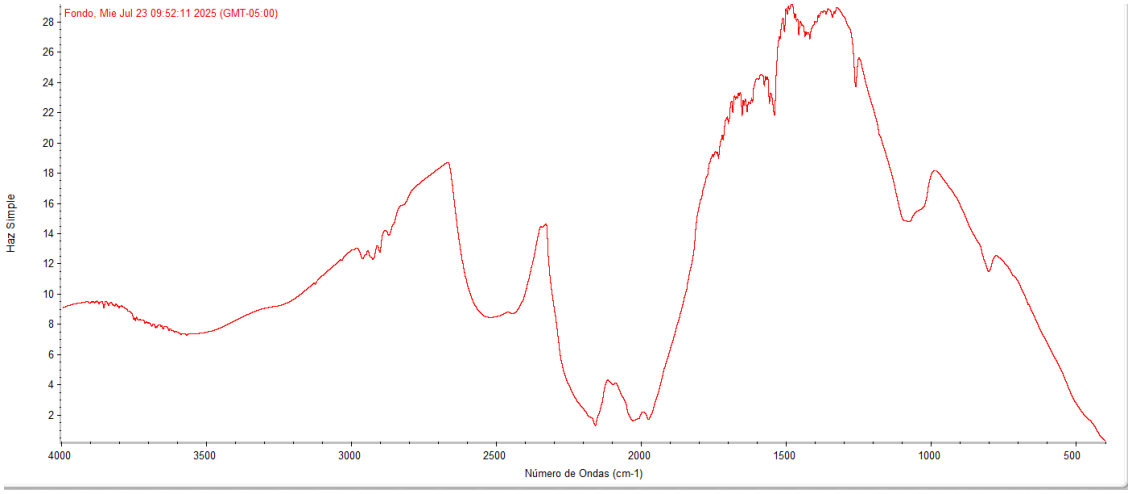


Figura 4: Background laboratorio de cromatografía .

Imagen elaborada por el Autor.

2.2.4 Espectro Etanol

Acorde al ejemplo presente se puede observar que los picos relevantes dentro del espectro del etanol al 96% marca los picos representativos entre 1047 cm-1, 1087 cm-1, 857 cm-1.

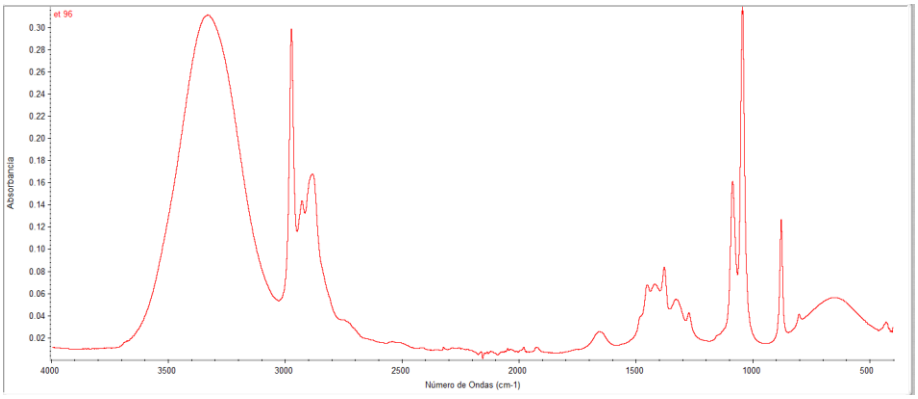


Figura 5: Espectro etanol al 96 º obtenido a través del equipo de FTIR.

Imagen elaborada por el Autor.

2.2.5 Espectro Metanol

En el caso del metanol los picos más representativos de sus enlaces se ubican entre los 1000 cm-1 y 1100 cm-1.

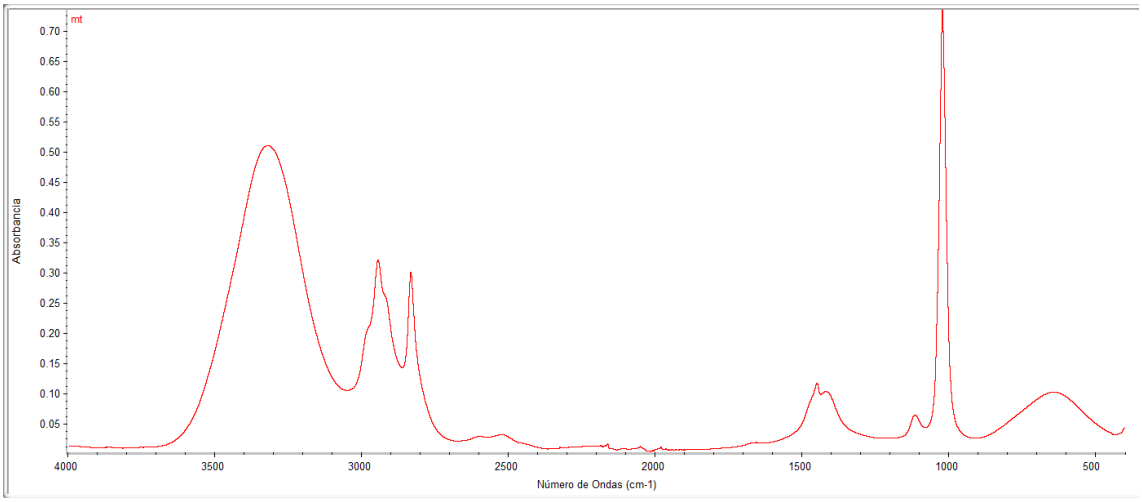


figura 6: Espectro de metanol absoluto obtenido a través del equipo de FTIR.

Imagen elaborada por el Autor.

2.2.6 Técnica ATR-FTIR

El espectrómetro de transformada de Fourier (FT-IR) es una herramienta analítica, cuyo objetivo es obtener un espectro que es obtenido cuando una muestra es irradiada simultáneamente bajo un amplo rango de frecuencias de espectro IR y las señales obtenidas pasa por el tratamiento y se les realiza la aplicación de la transformada de Fourier y así se descompone para identificar solo las frecuencias que fueron absorbidas. La técnica de reflectancia total atenuada o (ATR), es una modalidad de espectroscopia que su uso tiene ventajas como la rapidez, la facilidad de uso, mínima cantidad de muestra, pero sus ventajas también traen varias complicaciones como: número de onda, índice de refracción, profundidad de penetración de la onda evanescente, ángulo de incidencia, eficiencia de contacto, material de cristal, número de reflexiones.(Berthomieu & Hienerwadel, 2009).

2.3 Marco Conceptual

2.3.1 Pre Tratamientos matemáticos de datos espectrales.

2.3.1.1 Corrección de la línea base por medio de ALS

Al obtener los espectros crudos de equipo de FTIR estos tienen muchas señales añadidas conocidas como ruido, por lo que el primer pretratamiento consiste en la corrección por línea base esto con la finalidad de eliminar la mayor cantidad de ruido. Para corregir la línea base de nuestros espectros se debe eliminar las tendencias no deseadas. Para realizar esta corrección existen múltiples metodologías para conseguirlo.

Entre ellas figuran la corrección lineal, polinómica, por corrección de morfología y , ALS Asymetric Least Square, que va a usar en este estudio.

Para comprender la corrección por línea base, hay que tener en cuenta en que se trata de una función matemática en la cual se busca eliminarlas señales que no forman parte del espectro con el fin de aislar solo las señales analizables, mediante el método ALS se puede ajustar que tanto desea suavizar la línea, es decir eliminar las irregularidades, también se puede ajustar que tanto se acerca la línea base a los puntos inferiores del espectro y es asimétrica porque al momento de trazar la línea con las partes bajas del espectro es más relajado, en cambio cuando existen picos es más estricto.

ALS entonces es una fórmula en la cual penaliza dos situaciones: cuando la línea generada está por debajo de la línea base y cuando esta se encuentra entre los picos. Siendo más estricto cuando la línea se encuentra entre los picos, para ajustar esta función se controla a través de dos parámetros: el primero es lambda que esta controla que tanta suavidad se desea en el espectro, entre más grande este parámetro más se reducen los picos. Y el parámetro p en el que se asigna valores entre 0 y 1, en caso que se le asigne un valor de $p \approx 0$ va ajustar la línea base hacia a favor de los valles y va a darle menos peso los picos, el caso de que $p \approx 1$ realiza el caso contrario, es decir la línea base tenderá a seguir a favor de los picos menos peso a los valles y si es un punto medio como es $p \approx 0.5$ se realizará una línea base simétrica (Two Methods for Baseline Correction of Spectral Data • NIRPY Research, n.d.). esta función tiene ventajas como es la fiabilidad que genera, aunque el espectro tenga mucho ruido, o picos complejos va a generar una buena línea base, así mismo permite el ajuste de parámetros enfocado a lo que se requiere investigar si picos o valles lo que favorece para su aplicación para distintas áreas en espectroscopia, cromatografía, electrónica y electrónica.

Al usar este método en nuestro caso nos genera una gran ventaja la cual es la capacidad de eliminar picos anchos no significativos como lo es el pico del H2O que aparece en la número de onda de entre 3000 cm-1 a 4000 cm-1 en los espectros de metanol -etanol. El método ALS responde a la siguiente ecuación.

$$z = min \sum_{i=1} \quad W_i(y_i - z_i)^2 + \lambda \sum_{i} \quad (\Delta^2 z_i)^2$$

En donde:

y: Espectro

W_i: Son pesos asimétricos que depende entre la relación de "y" y "z"

λ: Es un parámetro de regularización sobre la suavidad permitida de la línea base.

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

2.3.1.2 Normalización de datos

Se realizó una normalización por área, esto tiene el propósito de que todos los datos se iguales entre sí, así que todos los datos cumplan los mismos requisitos, este caso que todos los datos tenga una mima área total., así mismo mediante ese método se tiene la capacidad de eliminar el sesgo en los datos y es un paso fundamental la preparación de los datos, debió a que hace que todos los datos estén en el mismo idioma por así decirlo Para esto ello se calcula con la siguiente fórmula.(RichmanJeffrey,n.d.).

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

En donde:

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

A_{λ}Corresponde a la absorbancia dentro de una onda número específica.

$\sum$ A_{λ}: Es la integral del espectro.

2.3.1.3 Suavizado del espectro y eliminación de ruido

Para suavizar señales y calcular derivadas se emplea el filtro Savitzky-Golay se lo puede describir como un filtro matemático de espectros que tiene la capacidad de suavizar datos ruidosos mediante mínimos cuadrados a el conjunto de puntos que se encuentran en puntos cercanos, además tiene la principal característica que mantiene la forma de los picos y los valles. Se puede describir el proceso en 3 pasos en donde primero se va a definir una ventana de puntos fijo y estos van a desplazarse en todos los datos, a partir de la ventana se genera una regresión línea y sitúas un nuevo punto central con el generado a partir de la regresión lineal.

Al aumentar estos valores existe un mejor suavizado, pero se corre el riesgo de que se pierdan detalles relevantes, al caso de disminuir la ventana de puntos fijo se tendrá mayor nitidez o mayores detalles, pero va a existe una menor eliminación del ruido, así mismo se puede interactuar con el grado del polinomio formado, entre mayor el grado va a existir una mejora dentro de las formas complejas formadas por el espectro, pero con el riesgo de sobreajuste el espectro. Dentro de la ecuación se tiene el parámetro los coeficientes de convolución estos son una operación matemática la cual permite aplicar un filtro formado por dos funciones que modifican una señal. En filtro se basa en la siguiente ecuación (Savitzky & Golay, 1964):

$$y_i^{(s)} = \sum_{j=-m}^{m} ciyi + j$$

En donde:

$y_i^{(s)}$: Representa el valor ya suavizado y la i señala de qué posición está ubicado.

$yi + j$: Responde a cuáles son los datos originales.

ci: Los coeficientes de convolución.

m : número de puntos vecinos durante el ajuste.

2.3.1.4 Primera y Segunda derivada

Para obtener la primera y la segunda derivada se puede hacer a través del mismo filtro S-G en el cual solo se modifica el grado del polinomio para obtener la primera o la segunda. La primera derivada tiene la capacidad de medir el cambio frente a intensidad en el eje x es principalmente útil para detectar y relatar borde en lo picos en cambio la segunda derivada sirve para representar la curvatura del espectro o en otras palabras la pendiente, se centra más en la inflexión o el cambio del espectro, es principalmente usado para separar señales espectrales que se encuentren muy cerca, también tiene la capacidad de evidenciar picos solapados, ayuda a eliminar el fondo que puede cambiar a la interpretación de los picos también se juega con la variable de coeficientes de convolución el cual ajustando el grado del polinomio se genera coeficientes de convolución específicos para la segunda y primera derivada. El filtro responde la siguiente ecuación. (Savitzky & Golay, 1964).

$$y_i^{(s)} = \sum_{j=-m}^{m} ciyi + j$$

Donde:

$y_i^{(s)}$: Representa el valor ya suavizado y la i señala de qué posición está ubicado.

$yi + j$: son los datos originales.

ci: los coeficientes de convolución.

m: número de puntos vecinos durante el ajuste.

2.4 Modelamiento matemático

2.4.1 Regresión Multidimensional

La regresión multidimensional es una técnica estadística que se puede considerar la evolución de la regresión lineal dado interrelación la variable independientes y dependientes es decir se quiere ver el cambio de las variables cuando una actúa en función de las variables predictoras(Granados, n.d.).

2.4.2 Mínimos cuadrados parciales (PLS)

La Regresión de Mínimos Cuadrados Parciales (PLS) es una técnica de regresión supervisada que tiene una cierta similitud con (PCA), esta tiene como objetivo ver la mayor varianza posible pero solo con los datos x , principalmente útil para la reducción la dimensional, en cambio el modelo PLS busca la máxima covarianza, es decir que los datos de X y Y se encuentran altamente relacionados. En otras palabras, busca generar una función que tenga la capacidad de predecir solo unas variables de respuesta (Y), con ayuda de un gran conjunto de variables independientes X, gracias a su naturaleza también permite ver la correlación de las variables X, que va busca que los componentes

maximicen las correlaciones. Para poder emplear este modelamiento se requiere tener todos los datos estandarizados para eso se realiza el proceso de visto previamente, una vez con todos los datos estandarizados se pasa por la extracción de los componentes y mediante la generación de un vector con capacidad de otorga pesos que maximizan la correlación entre las variables X y las Y a partir de la matriz se obtiene un score , el vector con la variable original se relaciona con los componentes latente para a partir de ahí realizar la predicción.

El modelo utiliza datos a los que se les nombra como train y corresponde como 70 % de datos estos nos van a servir para poder conocer el manejo de los datos y entrenar al modelo para luego evaluar con los datos de test que corresponde al 30 % con el fin de ver el rendimiento con datos nuevos, y va a seguir una limitación en la que solo va a predecir correctamente dentro del rango de los datos iniciales MSE. (Lê Cao et al., 2011). El modelado PLS se expresa mediante la siguiente ecuación y este mismo busca fragmentar las matrices en componentes latentes. Los loading son coeficientes que indican cuánto contribuye cada una de las variables en un componente latente. (Chun & Keleş, 2010).

$$X = TP^T + E$$

$$Y = UQ^T + E$$

En donde:

T= Matriz scores de componentes X

P = Loading de X

U = Matriz de con scores de Y

Q = Loadings de Y

$E\ y\ F$= Representa el error o residuo

Para el cálculo de las matrices se usa la siguiente fórmula en donde w son los pesos otorgados y x representa las variables originales de los componentes.

$$T = XW$$

Y para la predicción del modelo primero se debe obtener B que representa la matriz de regresión que son los coeficientes "X" y "Y" interrelacionados y se usa la siguiente ecuación

$$B = W(P^T W)^{-1} Q^T$$

Para la predicción se usa esta ecuación

$$Y = X * B$$

En donde y va a ser la variable de respuesta, x va ser el dato de entrada y mientras que b va a ser la matriz con los coeficientes para la predicción.

2.4.3 Sparse PLS

Es una variación o mejora del método PLS es especialmente bueno para ser utilizado en problemas con alta dimensionalidad, dado que su nombre indica que es un PLS pero con dispersó y eso indica con selección de variables. Eso lo hace cuando se tiene muchos datos y no todos aportan a la para la predicción entonces se seleccionan las variables más representativas para la selección se el Lasso lo que fuerza que algunos coeficientes se vuelvan cero. Y su principal objetivo es maximizar la covarianza entre X y Y, al forzar que algunos coeficientes se vuelvan cero contribuyen en menos sobreajuste y una mayor interpretabilidad. (Hyonho Chun & Sündüz Keleş, 2010).

De ahí el modelo procede igual que el PLS se entrena el modelo con el 70 % de los datos a los que se los conoce como train , y ya entrenado al modelo se lo enfrenta con los datos nuevos de datos test para evaluar su el comportamiento del modelo, gracias a la

gran capacidad de análisis de datos con alta dimensionalidad es especialmente útil en distintas ramas como son la proteómica y genómica (Fisher, 1936).

Responde a la siguiente ecuación en la cual busca la covarianza de los vectores w y c que busca en nuevas dimensiones los dos componentes latentes X, Y

$$MAX_{W,C} = Cova(X_{W,}Y_C)$$

Que esté sujeto a las siguientes matrices en las cuales representa S1 Y S2 el número de variables seleccionadas en X ; Y.

$$\|W\|_1 = 1\ PESO\ NORMALIZADOS$$

$$\|W\|_2 < S1 \quad PENALIZACIÓN$$

$$\|C\|_1 = 1\ PESO\ NORMALIZADOS$$

$$\|C\|_2 < S1 \quad PENALIZACIÓN$$

Para descomponer las componentes X y Y en donde T corresponde a la matriz de scores, en cambio P, Q son matrices de loading de X y Y

$$X = TP^T$$

$$Y = TQ^T$$

La fórmula en la que se basa la predicción es igual a la del modelo pls

$$B = W(P^TW)^{-1}Q^T$$

2.4.4 Tuneo de hiper parámetros

Consiste en un proceso de ajuste e optimización de hiper parámetros con el principal objetivo de que se obtengan los mejores hiper parámetros con el fin de mejorar el aprendizaje automático para la predicción además al optimizar los parámetros también ayuda a optimizar los códigos y el tiempo de ejecución, y para los modelos ayuda a reducir el sobre ajuste del modelo mejorar su eficiencia y velocidad de respuesta, minimizando los requerimientos computacionales. para la validación validar que si son las mejores señales para predecir se debe pasar por proceso de validación como lo es MSE. (APRENDE ESTADÍSTICAS FÁCILMENTE, n.d.).

2.4.5 Grid search

Es una herramienta de machine learning que consiste en otra metodología específica para el tuneo de hiper parámetros en la cual se realiza una búsqueda en forma de malla en el cual primero se define una rejilla en la cual contiene las posibles combinaciones, luego se prueban todas las combinaciones posibles y mediante un modelo de validación cruzada es posible evaluar el rendimiento de modelo.

Es completamente necesario dado que los modelos como el sPLS, dado que con unos buenos hiper parámetros aumenta la veracidad del modelo .(*Random Search and Grid Search for Function Optimization - MachineLearningMastery.Com*, n.d.).

2.4.6 Error Cuadrático Medio (MSE)

Permite evaluar la precisión y el rendimiento de los modelos generados, funciona a parir de una ecuación en la cual busca cuantificar la diferencia promedio entre los

valores obtenidos por el modelo y los valores reales de mismo, así obtenemos una media que describe la capacidad del modelo capacidad del modelo entre más bajo es el valor de RMSE este indica un menor error, o una correcta predicción de los valores predichos y los valores reales, Esta métrica responde a la siguiente ecuación (Christin Brettschneider etal.,2022):

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

En donde:

y_i: Representa a los valores reales del que se forma el modelo

$\hat{y}_i$: Representa los valores que fueron predichos gracias al modelo

n: Corresponde al número total de muestras.

2.4.7 Definición de términos

- **Transformada de Fourier**: Descripción de cómo se convierte la información de tiempo en el dominio de frecuencia para obtener el espectro.
- **Espectro FTIR**: Son las señales que fueron obtenidas a partir de 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.
- **Mínimos Cuadrados Parciales Disperso (sPLS):**se considera una extensión que parte de PLS en el cual él se selecciona las variables más representativas, esto permite que minimizar el uso de las variables irrelevantes, y para finalizar se hace una regresión sobre el mismo.

Capítulo 3

3.1 Materiales y métodos

3.1.1 Nivel de investigación

La presente investigación se clasifica como una investigación explicativa dado que busca encontrar la explicación entre la relación del alcohol de origen artesanal y concentración de metanol-etanol presente en la muestra, y acorde al pico mostrado por espectroscopia crear un modelo que prediga si existe o no presencia de metanol.

3.1.2 Diseño de investigación

Dentro del presente documento se desarrollará un diseño de investigación experimental, ya que se interactúa con las muestras con distintas concentraciones de Etanol-Metanol se someten para el análisis y elaboración de un modelo predictivo. Para el diseño de investigación se hizo a través de la técnica de análisis de contenido con el fin de obtener las partes de interés de cada estudio, esto se hizo gracias a la investigación de dentro de bases de datos como lo son SCOPUS, GOOGLE ACAEMICO, LATINDEX y Repositorio de universidad UPS entre otras.

3.1.3 Diseño de experimento

Para el desarrollo del modelo estadístico es necesario generar nuestra base de datos espectrales, el objetivo es darle un enfoque en el cual se permite crear una librería de estandartes representativos y que tengan variabilidad en las distintas concentraciones de Etanol y Metanol, que se ajusten a concentraciones en contextos reales.

Para ello se opta por un diseño de mezclas que es expresada

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

xi Proporción del componente i (metanol-etanol)

lc Límite inferior permitido para el componente i

uc Límite superior permitido para el componente i

$0\ y\ 1$ Rango proporcional

Para esta investigación se sitúan los rangos de las concentraciones del alcohol en

Codificado $0 < \text{Etanol} < 1$

No Codificado $30 < \text{Etanol} < 50$

Codificado $0 < \text{Metanol} < 1$

No Codificado $0 < \text{Metanol} < 2$

Estos rangos son los de concentraciones en los que se va a trabajar dado que en el caso del etanol estos rangos están dentro de el volumen de alcohol que normalmente deben según las normas INEN para este tipo de bebidas, en caso del metanol se estableció el rango entre 0 % y 2% dado que acorde a la norma el mismo no ser mayor de 0.2 además al tener múltiples concentraciones permite tener un espectro más detallado que nos ayuda al análisis de los datos, en total se generaron alrededor de 72 concentraciones.

3.1.4 Técnicas e instrumentos de recolección de datos

Para este estudio se emplearán distintas técnicas para la recolección de datos entre las cuales figuran están como punto de partida, la elaboración de los estándares, el tratamiento de cada estándar,para ello se utilizo observación instrumental la cual consistió en la medición espectral mediante el equipo, el registro de cada espectro crudo fue en formato CVS., los instrumentos empleados para obtener los espectros fueron el equipo de espectroscopia FTIR, software "OMNIC" para la visualización y obtención de los espectros.

3.1.5 Técnicas de procesamiento y análisis de datos

Para las técnicas del procesamiento de datos se van a emplear varias técnicas matemáticas en las cuales figura la corrección de línea base (ALS), suavización del espectro, normalización de escala, primera y segunda derivada por el filtro Savitzky-Golay, para el análisis de cada una de las variantes, las técnicas utilizadas para la formación de los modelos fueron elaborados de modelos de regresión multilínea fueron el sPLS y PLS. Para el análisis de los espectros y el modelamiento de los modelos el software empleado fue de R- studio.

3.1.6 Variables

3.1.6.1 Variable independiente

Corresponde a las distintas longitudes de onda de cada una de las distintas concentraciones de metanol y etanol de los espectros analizados, en el cual vamos a recibir a manera de feedback las distintas absorbancias

3.1.6.2 Variables dependientes

Las variables dependientes corresponden a las distintas Concentración de Etanol-Metanol para la formación del modelo predictivo.

3.1.6.3 Variables Intervenes

Van a representar a los factores que afectan la relación entre las variables independientes y las dependientes entre ellas figura, la temperatura durante el análisis, Condiciones de preparación de la muestra. El volumen de la muestra, y la experiencia del analista.

3.3 Metodología

3.3.1 Fase preexperimental

Si bien la fase preexperimental parece sencilla, es el primer paso para la formación de los distintivos espectros para la posterior formación del modelo matemático, y su principal objetivo de esta sesión tiene como objetivo eliminar la posible variabilidad que puede afectar a los espectros con el fin de que el presente trabajo pueda ser repetible y su futuro uso en trabajos posteriores.

Para verificar la veracidad de las muestras espectrales se realizó 3 corridas con etanol al 35% con el fin de ver las posibles variaciones en los espectros con el fin de ver la variación que el equipo posee, acorde a la figura inferior se observó una variabilidad prominente de la misma muestra por lo que es algo a tomar en cuenta para futuras investigaciones.

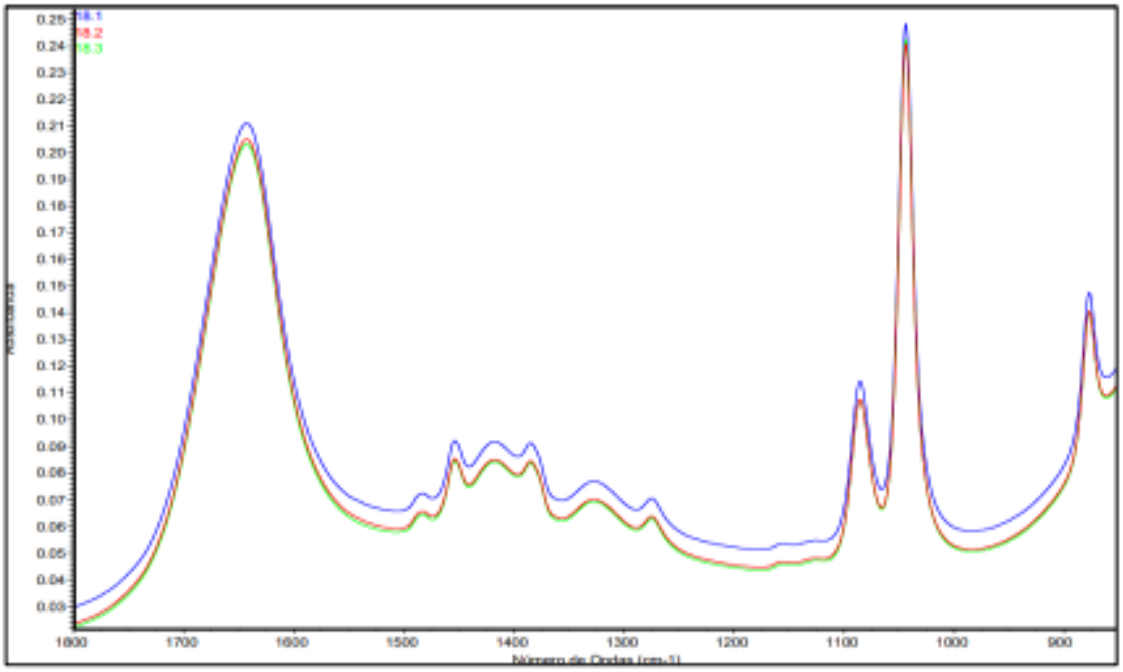


figura 7: Varianza de la medición de equipo de una misma muestra de alcohol.

Imagen elaborada por el Autor.

3.3.2 Efecto de la muestra

Acorde y según trabajos anteriores de espectrofotometría se investigó que acorde a al volumen de la muestra varía las absorbancias de los aspectos y se observó que cuando el volumen de la muestra sobrepasa los 55 microlitros no existe cambios por lo que sugiere que este ese sería el volumen optimo.

Además, se observó que dado el lugar y las características del laboratorio se puede inferir que existe una alta humedad relativa lo que infiere en la evaporación del etanol sobre la punta del analizador esto genera en el patrón un aumento de la absorbancia por lo que gracias a trabajos anteriores que se estableció que el tiempo de evaporación del etanol para que no existan aumentos en la absorbancia es de dos minutos y se debe realizar una limpieza con agua destilada en la superficie del máquina.

Para la preparación de cada estándar para la formación de los espectros deben asegurar una mezcla totalmente homogénea por lo que cada muestra se pasó por un proceso de agitación de vórtice durante 15 segundos.

Dependiendo de las concentraciones de los estándares espectrales se hacen distintas disoluciones en relación con la cantidad de etanol-metanol en distintas proporciones, dado que cada uno va a delimitar un dato que va a servir para nuestro modelo, después con la ayuda de un agitador magnético se homogeniza bien, una vez bien establecido las muestras y las diferentes concentraciones se pasa para el análisis en el equipo FTIR en donde se pasaran las muestras y así obtendrá los distintos espectros que son analizados y pasan a datos que luego se transforman en bases de datos en "R studio" donde los datos serán tratados por los modelos estadísticos antes mencionados para poder evaluar los distintos modelos matemáticos y pueden predecir la cantidad de metanol-etanol de la muestra. Y para finalizar comparar con la muestra real la cual antes se debe pasar por cromatografía de gases para ver su composición evaluar si el modelo predice la concentración o no.

3.3.3 Consideraciones del equipo

En el momento de analizar cada muestra dentro del equipo FTIR con el software de OMNIC este mismo establecer un número de scans o análisis de entre 3 a 42 esto quiere decir el equipo recoge analiza y promedia la señal recibida, al aumentar el número de scans aumenta la calidad de la señal y disminuye el ruido en la señal, pero en consecuencia también el tiempo de toma de muestra, por lo que un número equilibrado entre tiempo y calidad de la señal estaría entre unos 20 a 32 scans.

Parámetros experimentales para el uso de equipo:

- Volumen de muestra: 55µl

- Tiempo de espera entre muestra: 2 min
- Tiempo de vórtex: 15 seg
- Resolución espectral: 4 $cm{-}1$
- Espaciado de puntos: 0,4821 $cm{-}1$
- número de scans: 32
- Background: un background por cada réplica y diferente muestra

3.4 Fase experimental

3.4.1 Preparación de los estándares

El primer paso de la fase experimental consistió en la preparación de los estándares para ello se emplearon tubos Eppendorf de 2 ml y con ayuda de una micropipeta se preparó 1 ml de muestra para cada concentración en los distintos niveles y luego pasó por su posterior etiquetado. Para evitar errores para la elaboración se hicieron dilaciones volumen- volumen., en caso del metanol al ser muestras pequeñas se optó por realizar una solución madre al 5 % con agua milli-Q y fue usada en cada uno de los distintos niveles. Para la elaboración de cada uno de los niveles se usó la siguiente tabla.

NIVEL 1	ETANOL	30%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00
NIVEL 2	ETANOL	32.5%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00
NIVEL 3	ETANOL	35%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00
NIVEL 4	ETANOL	37.5%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00
NIVEL 5	ETANOL	40%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00
NIVEL 6	ETANOL	42.5%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00
NIVEL 7	ETANOL	45%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00
NIVEL 8	ETANOL	47.5%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00
NIVEL 9	ETANOL	50%							
	METANOL	0.00	0.05	0.10	0.30	0.50	1.00	1.50	2.00

Figura 8: Tabla de mezclas de para la elaboración de los distintos estándares.

Imagen elaborada por el Autor.

3.4.2 Selección del rango

En trabajos el presente trabajo se buscó comparar cuál de los modelos da una mejor predicción. Al momento de construir los modelos a partir de los espectros de los estándares se lo hizo con dos enfoques distintos, el primera consiste en usar todo el espectro para la construcción del modelo y el segundo enfoque se centra en limitar la número de onda dentro de los rangos de interés en los que aparece el etanol- metanol (800 cm¨-1 ;1250 cm-1), En otras palabras, el rango seleccionado para la primera para el primer modelo se usa todo el espectro todo el espectro y para el segundo solo se arma el modelo a partir de la región delimitada entre los 800 cm¨-1 ;1250 cm-1.

Los datos fueron obtenidos de manera cruda es decir sin ninguna modificación dentro del software OMNIC, del equipo los datos fueron guardados en formato .CSV este es el formato ideal para el posterior procesamiento de datos con ayuda de R studio.

3.4.3 Preprocesamiento de datos

Una vez obtenidos los múltiples espectros acorde a las distintas concentraciones de etanol y metanol se obtuvieron alrededor de 8 concentraciones de metanol para cada porcentaje de etanol y eso representaba un nivel de los nueve registrados, en total se obtuvieron alrededor de 72 archivos en formato CSV. Estos datos se encontraban en crudo es decir sin ningún tratamiento sino se refiere a que los datos obtenidos por el equipo fueron insertados en R Studio para el posterior análisis, para una correcta lectura

de los datos se emplearon dos paquetes *ir y dplyr*. Al normalizar los datos se hicieron a través de métodos de tratamiento de señales, entre ellos incluye línea base, normalización, suavizado y segunda derivada para los datos del espectro completo y de solo la parte delimitada.

Una vez ingresados los datos al se obtiene el siguiente espectro el cual representa en el eje x la número de onda y en el eje y las absorbancias de todas las muestras procesadas.

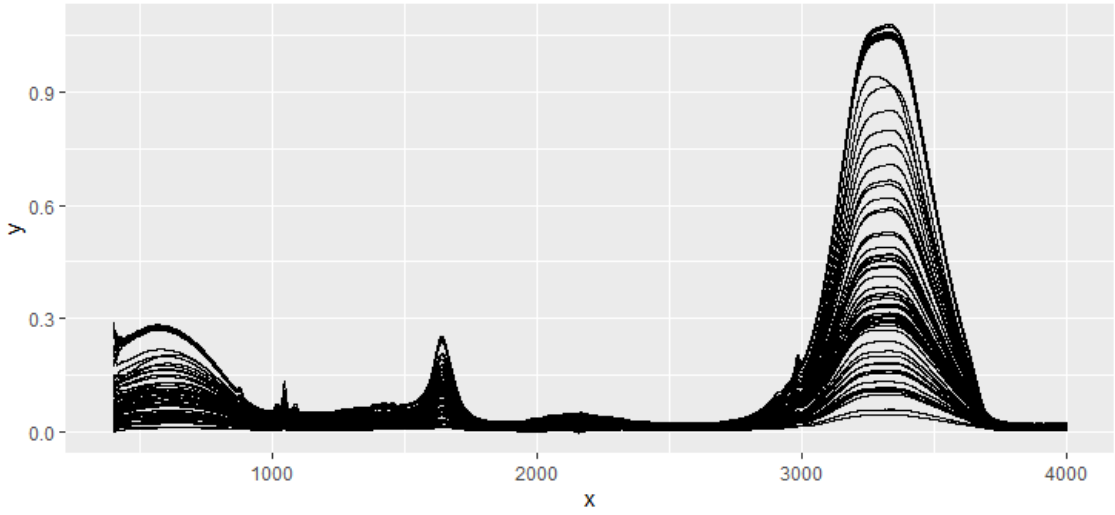


Figura 9: Todos los espectros de cada uno de los estándares.

Imagen elaborada por el Autor.

Una vez con los datos ingresados al programa de r estudio se optó por hacer dos scripts en el cual el primero fue tomando todo el espectro y en cambio el segundo se realizó un corte solo delimitando en los picos de interés y a partir de este tratamiento se realizaron cada una de las correcciones a los espectros.

3.4.4 Corrección de línea base

Se aplicó una corrección de línea base y tal como su nombre lo sugiere lo que se busca es que esta técnica elimina lo máximo posible el desplazamiento del fondo o también dicho eliminar las señales base no deseadas, es decir eliminar los picos no

representativos para que los datos sean representado de con mayor calidad y precisión, para hacerlo se escogió el método de ALS (Mínimos cuadrados asimétricos). En este algoritmo separa el espectro en dos partes, la señal analítica real y una línea base suave, se ajusta la línea base suave y está penaliza a las regiones que se salgan de por encima de curva de la señal analítica real, así se obtiene una línea base que no interfiere con los picos de la señal, además con el método ALS también realiza la eliminación del pico de agua en los 3000 cm-1 a 4000 cm-1 para este proceso en R-Studio se hace una función con ayuda del paquete **"baseline ".**

3.4.5 Suavizado del espectro

Para que el espectro pase por un proceso de suavizado el objetivo este tratamiento consiste en la reducción del ruido sin distorsionar el espectro ni la forma de los picos para eso se usó el filtro de Savitzky-Golay este método se basa en el cálculo de una regresión polinomial local (de grado k), con al menos k+1 puntos equiespaciados, para determinar el nuevo valor de cada punto. El resultado será una función similar a los datos de entrada, pero suavizada.el mismo que nos sirve para la segunda derivada, pero con la diferencia que en el filtro se coloca m=0.

3.4.6 Normalización de datos

Después de la corrección de línea base y del suavizado con los datos en el espectro se procedió al proceso de normalización. La función lo que hace es dividir cada uno de los valores de intensidad por suma de todos los valores de intensidad con el fin de estandarizar.

3.4.7 Segunda derivada

La segunda derivada fue calculada con mediante el filtro de Savitzky-Golay y se lo hizo junto al paquete" singal" esto permite la modificación de las variables p, n y m. dentro del análisis se usaron p = 2 que indica el grado de polinomio, n = 7 son el número de las variables de la número de onda que van a ser empleados para a la elaboración del polinomio, m = 2 que indica que sé que quiere calcular la segunda derivada.

3.5 Métodos estadísticos

3.5.1 PLS

Una vez con los datos pasados por el pretratamiento preliminar al siguiente paso fue la elaboración de los modelos para ello se elaboración dos modelos uno a través con el espectro completo y otro solo con los picos de interés por ello se empleó regresión por mimo cuadrados parciales para genera el modelo se empleó el paquete mixomics. Este paquete permite la ayuda a realizar la selección y clasificación en una sola acción.

3.5.1.1 Entrenamiento del modelo

Una vez generado el modelo con los datos son separados en dos conjuntos el 70 % de los datos van a ser empleado dentro del modelo para que el mismo se entrene la capacidad de predecir y en caso de el 30 % restante van a ser sometidos a la prueba del modelo y en respuesta vamos a obtener el margen de error que produce el modelo.

3.5.2 Sparse PLS

Es un complemento para el método PLS o una mejora del modelo que es utilizado para abordar problemas con alta dimensionalidad, funciona bajo un mecanismo que

permite identificar las variables más representativas para la predicción estas son escogidas mediante penalización así eliminado las menos importantes así que a partir de esta metodología se armaron igual que el PLS se armaron dos modelos uno con el espectro completo mientras que el otros solo con los picos de interés.

3.5.2.1 Entrenamiento del modelo

Y de mismo modo que con el PLS una vez el modelo ya estaba completo se procedió con el mismo método de entrenamiento en donde los datos fueron separados en dos conjuntos el 70 % de los mismo van a ser usados para el entrenamiento del modelo y para evaluar el modelo se tomó el 30 % de datos restante para ser probados en el modelo y la respuesta que vamos a obtener va a ser el de error que produce el modelo al momento de predecir.

3.5.3 Error Cuadrático Medio (MSE)

Una vez que el modelo fue armado, el siguiente proceso fue evaluar la funcionalidad del mismo para esto se realizó mediante el cálculo de Error Medio Cuadrático en las dos componentes principales, para esta evaluación se implementó una función dentro del código que proporciona una lista con valores en que representa el error que comete el modelo en cada componente, en resumen podemos decir que es una comparación entre los valores reales frente a los predichos y la diferencia que se genera luego es promediada.

3.5.4 Predicción con las muestras de alcohol artesanal de yunguilla

Al obtener el modelo, así mismo comprobando que proporciona un error cuadrático bajo, el paso final consiste en evaluar las cuatro réplicas del espectro producido por el alcohol artesanal de yunguilla y colocados dentro del modelo para que el modelo de la predicción la cantidad de etanol y metanol presente en la muestra.

Una vez con todos los datos ya tratados el paso final consiste en insertar el espectro tratado dentro del modelo para que entregue la predicción de las concentraciones de metanol-etanol.

3.5.5 Comprobación de la predicción del Alcohol Artesanal.

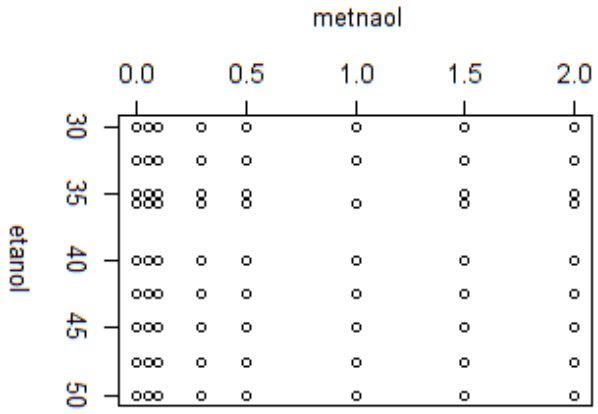
Para comprobar las concentraciones de metanol- etanol entregados por el modelo, se envió una muestra de aproximadamente de 300 ml hacia el laboratorio MVS el cual es una institución que ofrece servicios de análisis fisicoquímicos y microbiológicos para el control de calidad de alimentos y a través de técnicas de análisis la entidad va a dar las concentraciones exactas de metanol etanol, y en función a eso podemos efectividad del estudio.

Capítulo 4

4.1 Resultados y discusión

4.1.1 Determinación de la interrelación entre las concentraciones de etanol metanol y sus espectros.

Para establecer la interrelación entre las concentraciones de etanol-metanol y sus respectivos espectros infrarrojos, se analizaron los cambios de la en las señales espectrales en función de la variación de las concentraciones, Para llevarlo a cabo primero se elaboraron las concentraciones iniciales siguiendo los patrones de dilución que se muestra en la *figura 10* y cumpliendo con la metodología planteada en el capítulo 3, y se procedió al análisis mediante el equipo FTIR.

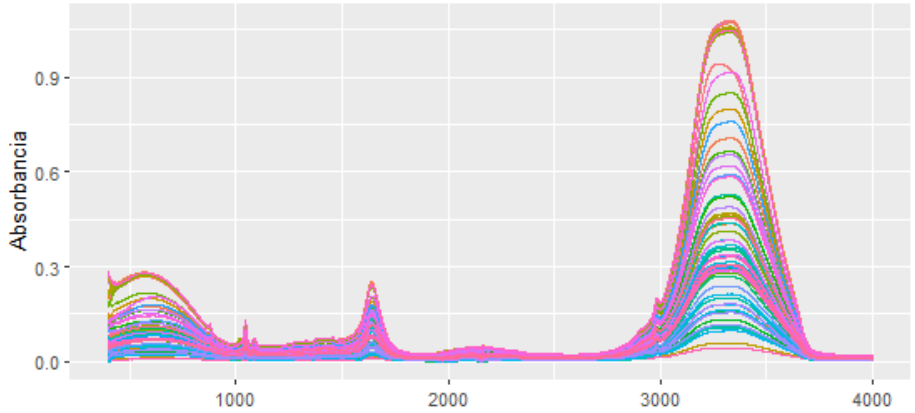


Figuras 10: En la figura se muestra las distintas concentraciones de metanol-etanol

Imágenes elaboradas por el autor.

Una vez con los espectros en formato en formato plano (CSV) se importó al software R-studio para su posterior análisis, Dentro de la figura 11 se observa el espectro formado por las distintas concentraciones y en el mismo se aprecia las variaciones entre las concentraciones de alcoholes presentes y las regiones específicas del espectro. Particularmente entre las bandas cercanas a los 3300 cm $^{-1}$ y 1000 cm $^{-1}$. Estas zonas están

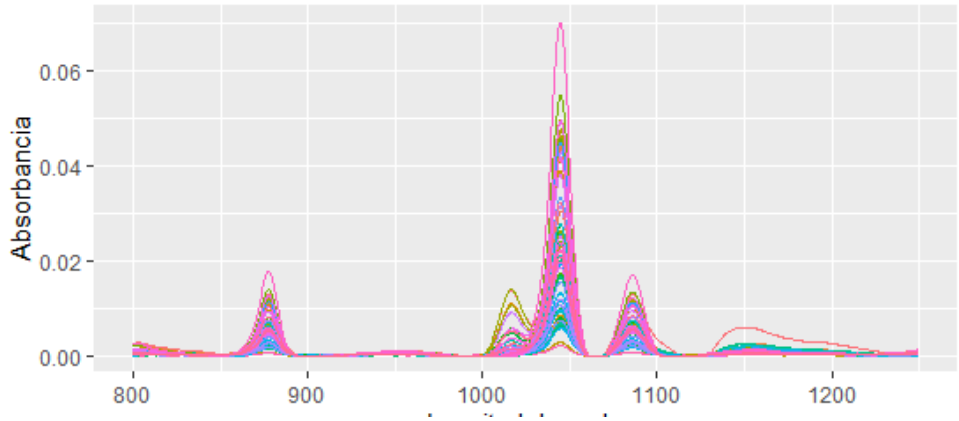
relacionadas con los grupos funcionales característicos del etanol y metanol, como los enlaces O–H y C–H, que presentan vibraciones distintivas y sensibles a los cambios de concentración.



e los 72 espectro leccionó el núme nientras que el segundo archivos de secuencia se lo analizó por completo con toda la número de onda 100cm-1 a los 4000 cm-1. Tal y como se indicó en el apartado de "Preprocesamiento de datos".

El primer tratamiento consistió en la corrección de la línea base a través del método de Mínimos cuadrados asimétricos y con ayuda del software r studio, cómo se aprecia dentro de la figura 11 y 12, se muestra la correlación positiva entre las distintas concentraciones, y con cada uno de los espectros generados, además se aprecia cómo cambia la absorbancia de cada concentración pero manteniendo la misma estructura de los picos en los mismos número de ondas, este método nos permite eliminar los picos de fondo y el pico correspondiente del agua,esto con relación de los archivos en la base de datos del espectro completo mientra que en el caso de segundo archivos se realiza el

mismo método para eliminar los picos de fondo considerando que en este archivos no está presente el pico.



Figuras 11 : Espectro obtenido luego de la corrección por línea base proviene de archivos que fue cortado

Espectro cortado

Imágenes elaboradas por el autor.

Espectro completo

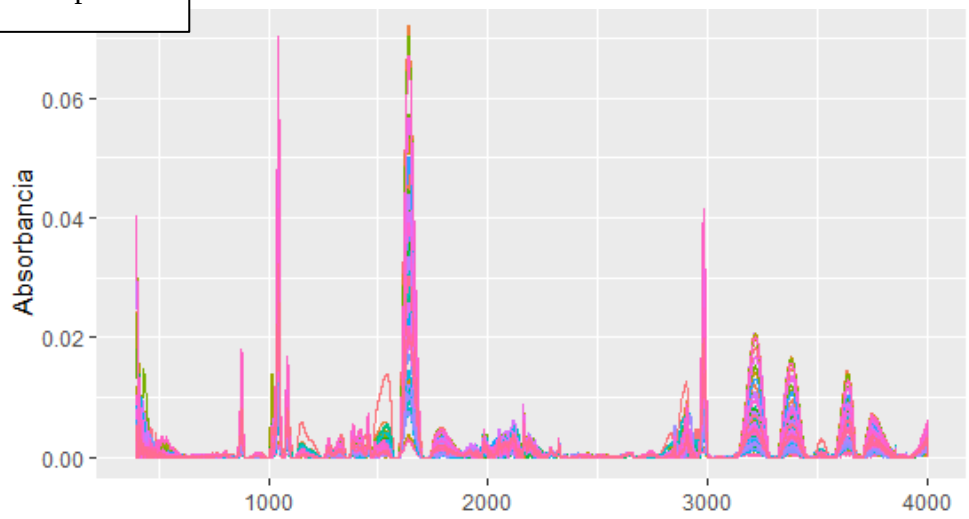
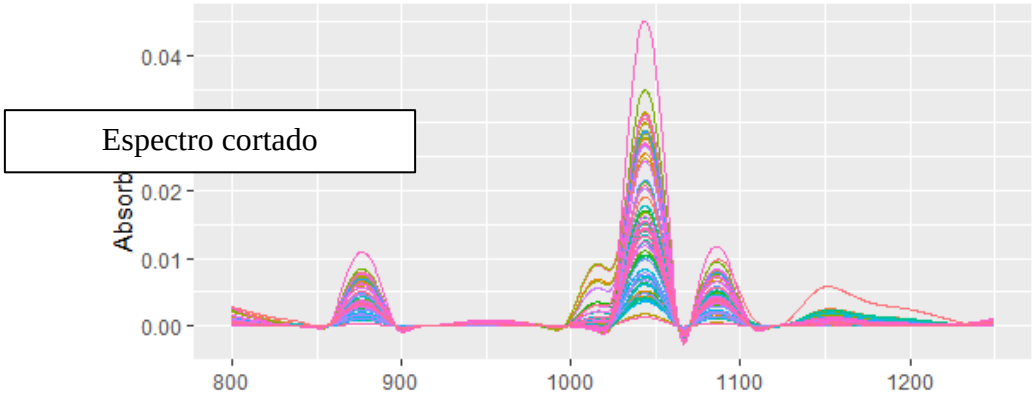


Figura 12: Espectro obtenido luego de la corrección por línea base del archivos tomando en cuenta el espectro completo.

Imágenes elaboradas por el autor.

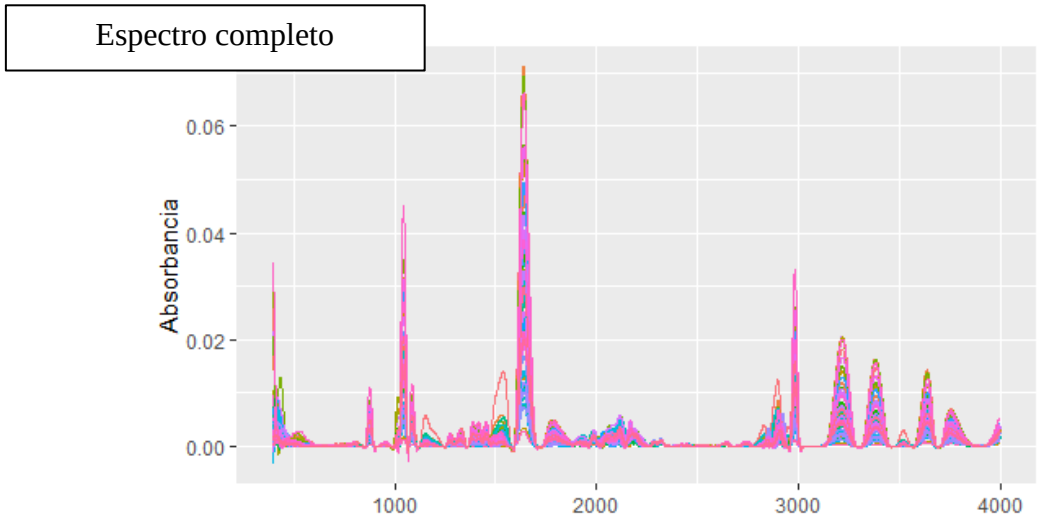
Después se procedió con el suavizado del espectro para eso se lo realiza mediante la técnica de suavizado (Savitzky-Golay), esta técnica se encarga de suavizar los espectros

mediante el ajuste de datos de pequeños tramos de la señal a polinomios como se aprecia dentro de la figura 13 y 14 . Así se redujo el ruido sin distorsionar la forma general de los datos, preservando bien picos o transiciones importantes.



Figuras 13 : Espectro obtenido luego del suavizado proviene de archivos que fueron cortados.

Imágenes elaboradas por el autor.

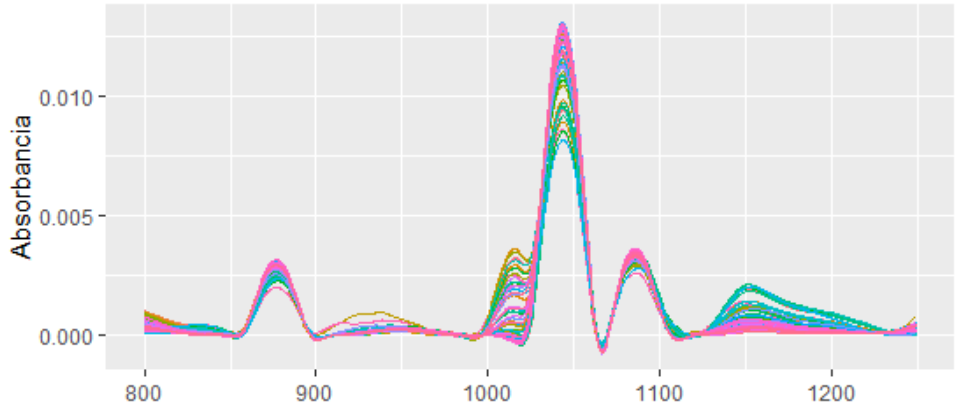


Figuras 14 : Espectro obtenido luego del suavizado de archivos tomando en cuenta el espectro completo.

Imágenes elaboradas por el autor.

Lo penúltimo fue la normalización del espectro. Este proceso se lo observa dentro del Figura (..)(..), consistió en dividir cada valor de intensidad por la suma total de intensidades del espectro. La finalidad de este método es estandarizar los datos, de modo que todos los espectros queden en la misma escala relativa.

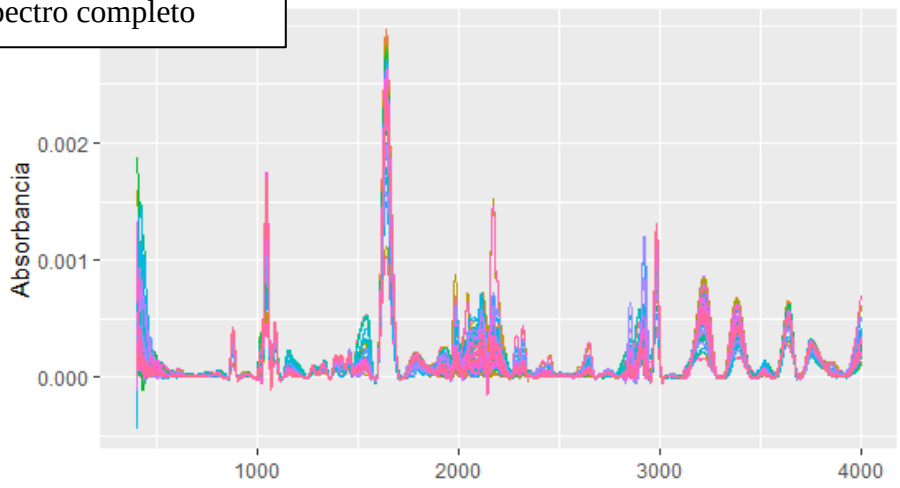
Espectro cortado



Figuras 15 : Espectro normalizado formado a partir de archivos en el cual es espectro fue cortado.

Imágenes elaboradas por el autor.

Espectro completo



Figuras 16 : Espectro normalizado formado a partir del espectro completo

Imágenes elaboradas por el autor.

El paso final consistió en la obtención de la primera y segunda derivada de los espectros, utilizando nuevamente la técnica de Savitzky-Golay. Esta técnica permite calcular derivadas desde el primer hasta el quinto orden, conservando la forma general del espectro mientras resalta sus características más importantes. Para su aplicación, se

utilizó el entorno de R Studio mediante la librería “signal”, que facilita la implementación del método de forma eficiente y precisa.

Como se aprecia dentro de la figura 17 y 18 , se puede observar la primera derivada permite el cálculo de la pendiente del espectro, de esta manera se resaltan los cambios en la intensidad.

Primera derivada

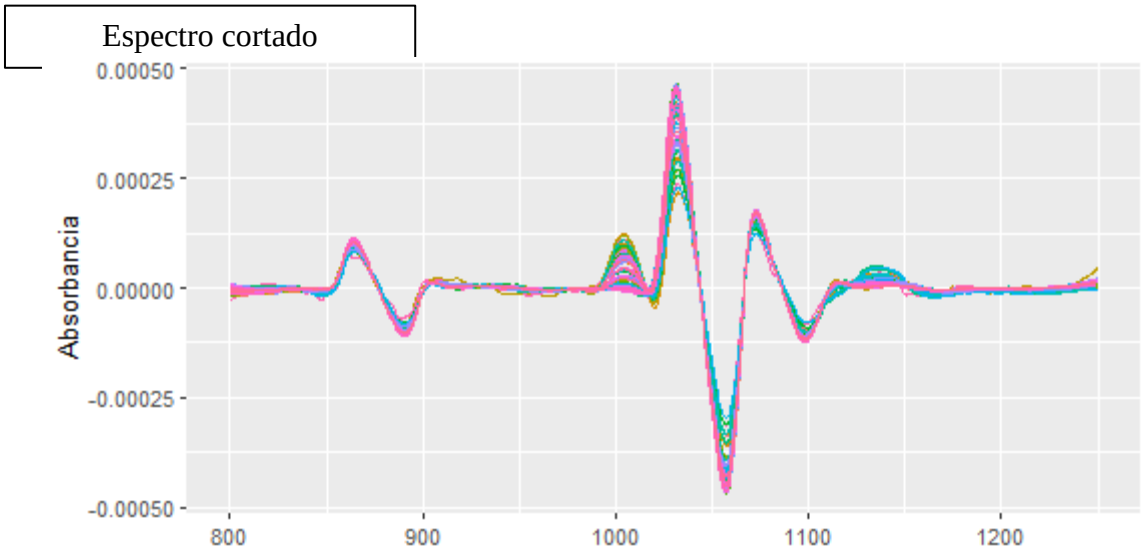
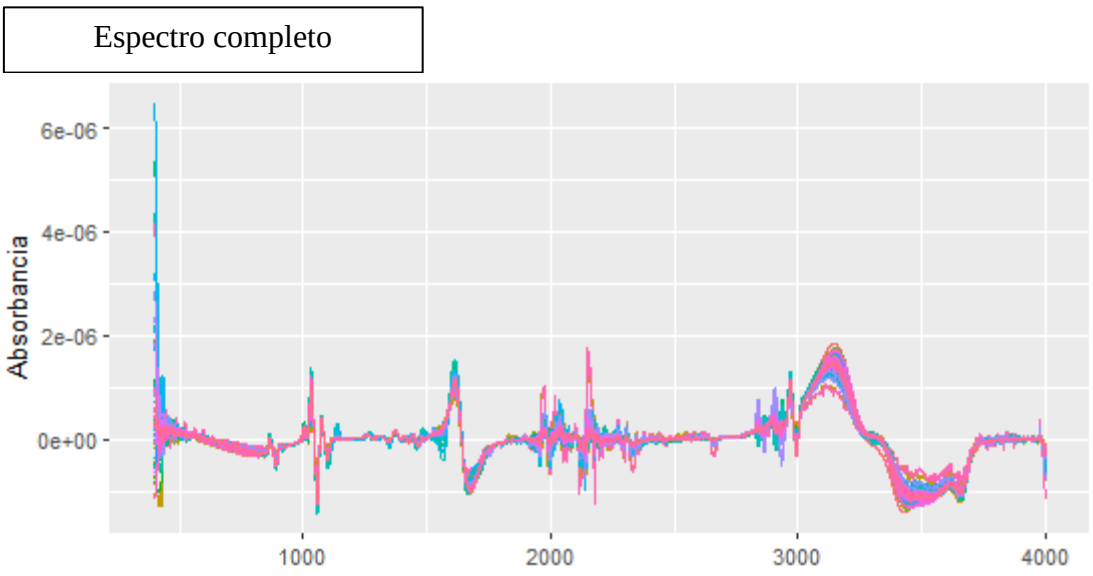


Figura 17: Espectro obtenido de la primera derivada formado a partir de archivos en el cual es espectro fue cortado.

Imágenes elaboradas por el autor.



Figuras 18 :Espectro obtenido de la primera derivada formado a partir de archivos en el cual es espectro completo.

Imágenes elaboradas por el autor

La segunda derivada esta permite el cálculo respecto a la curvatura del espectro. Resalta dónde hay un máximo o mínimo local, haciendo que los picos reales se vean más definidos.

Espectro cortado

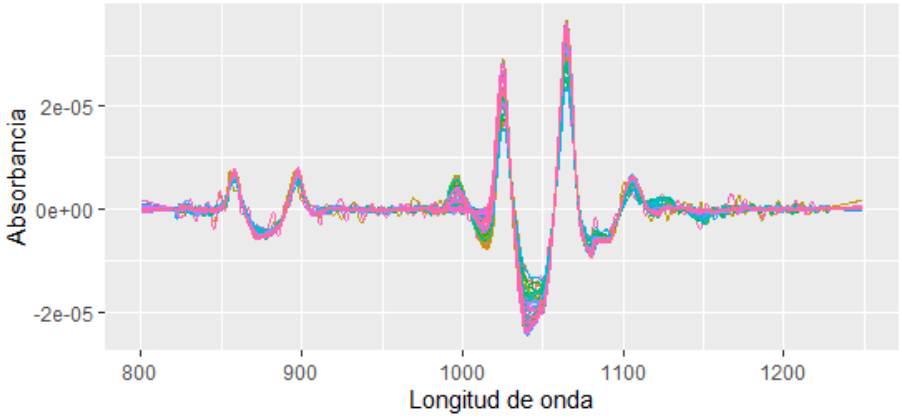


Figura 19: Espectro obtenido de la segunda derivada formado a partir de archivos con el espectro completo. *Imágenes elaboradas por el autor.*

Espectro completo

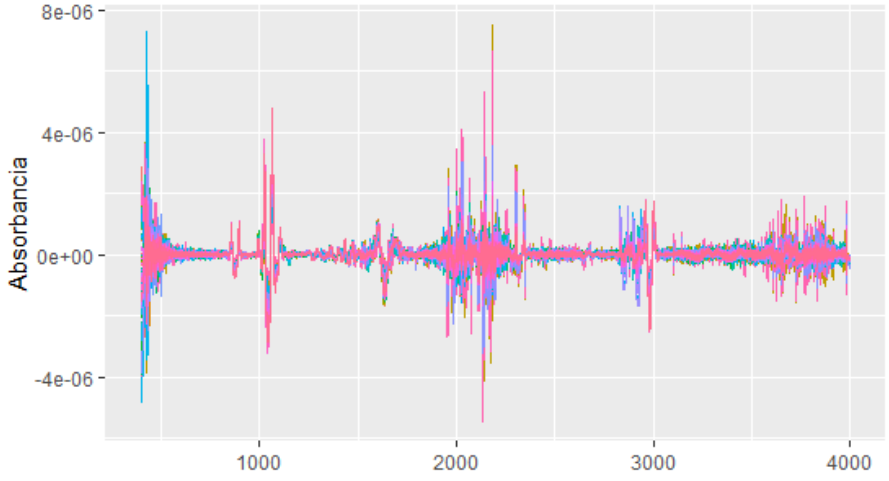


Figura 20 : Espectro obtenido de la segunda derivada formado a partir de archivos en el cual es espectro completo.

Imágenes elaboradas por el autor.

A partir de todos los espectros procesados, se logró observar cómo las distintas concentraciones se interrelacionaron con los espectros generados, evidenciando una correlación positiva que no se vio afectada por los distintos tratamientos aplicados a los datos.

4.1.2 Evaluación de modelos predictivos utilizando datos obtenidos de los espectros para realizar predicciones simultáneas de las concentraciones de etanol y metanol.

Partiendo de los datos procesados, se procedió a la elaboración del modelo de regresión multilineal. Para ello, fue necesario generar un tercer archivo.

A partir del archivo que contiene el espectro recortado y el espectro completo, se procedió a la elaboración de los modelos mediante la metodología de mínimos cuadrados dispersos, utilizando el software RStudio y el paquete "mixOmics". Esta metodología corresponde a una regresión multivariada que se enfoca en seleccionar únicamente las variables más representativas. En contraste, al emplear el modelo basado en mínimos cuadrados tradicional, se observó un sobreajuste del modelo debido al gran número de variables presentes en los espectros. Por lo tanto, la metodología de mínimos cuadrados dispersos, se consolidó como la opción más adecuada para este análisis.

Para garantizar la fiabilidad de los modelos, se aplicó un proceso de validación cruzada, dividiendo el conjunto de datos en segmentos de entrenamiento y prueba. Esto me permitió evaluar el rendimiento real del modelo al predecir nuevas concentraciones a partir de los espectros. Como métrica de evaluación se utilizó el Error Cuadrático Medio

(MSE), el cual permite cuantificar la precisión del modelo al comparar los valores predichos con los valores reales. Cuanto menor es el valor del Error Cuadrático Medio, mayor es la capacidad de predicción del modelo. En la tabla 1 se presenta el error cuadrático medio del modelo obtenido con la primera y la segunda derivada, relacionado al espectro cortado y al espectro completo.

	Espectro completo		*Espectro cortado*	
	Primera derivada	**Segunda derivada**	**Primera derivada**	**Segunda derivada**
	MSE	MSE	MSE	MSE
	0.55	0.37	0.64	0.54

Tabla 1: En la cual se muestran los diferentes MSE el cual permite valorar la confiabilidad del modelo.

Imágenes elaboradas por el autor.

De acuerdo con los resultados mostrados en la tabla 2 reveló que los dos grupos, espectro completo y el espectro cortado, con el menor error en la predicción simultánea de las concentraciones de metanol y etanol fue aquel que empleó la segunda derivada. Estos modelos alcanzaron un error medio cuadrático de 0,38 y 0,54 lo que indica una alta capacidad predictiva durante la validación cruzada. además En el caso contrario los peores modelos fueron generados a partir de la primera derivada.

4.1.3 Evaluación de los datos obtenidos del modelo en comparación con los valores reales presentes en la muestra.

Establecido el mejor modelo para la detección simultánea de los alcoholes el paso final fue la comparación entre las concentraciones obtenidas a través de modelo y los resultados obtenidos por en el laboratorio que se encuentra dentro de la tabla(). Para ese punto primero se tuvo que elaborar el mismo pretratamiento que se mencionó en capítulo anterior.

4.1.3.1 Pretratamiento de los espectros

Como punto de partida, se recolectaron cuatro muestras de alcohol artesanal. A partir de estas, se obtuvieron sus espectros. Posteriormente, los datos adquiridos se analizaron en R-Studio, generándose cuatro espectros, uno de los cuales se muestra en la Figura 22.

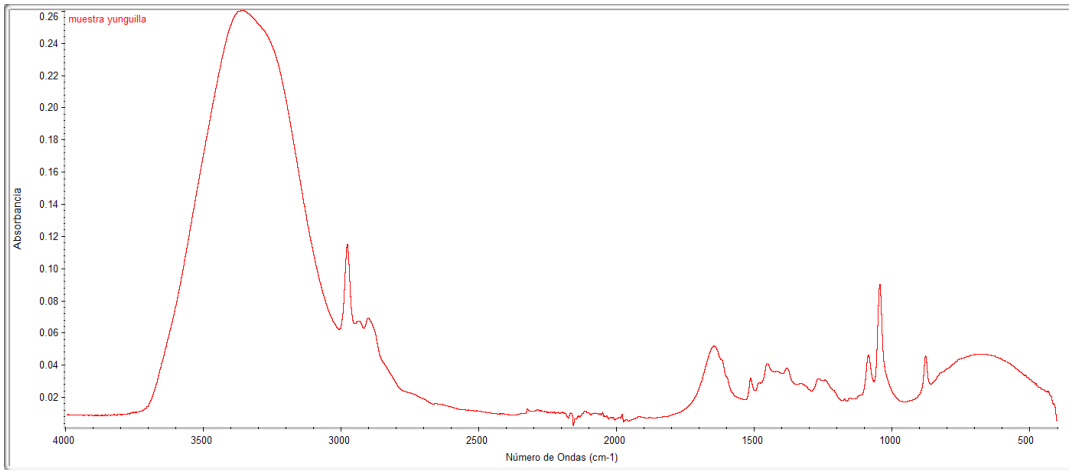
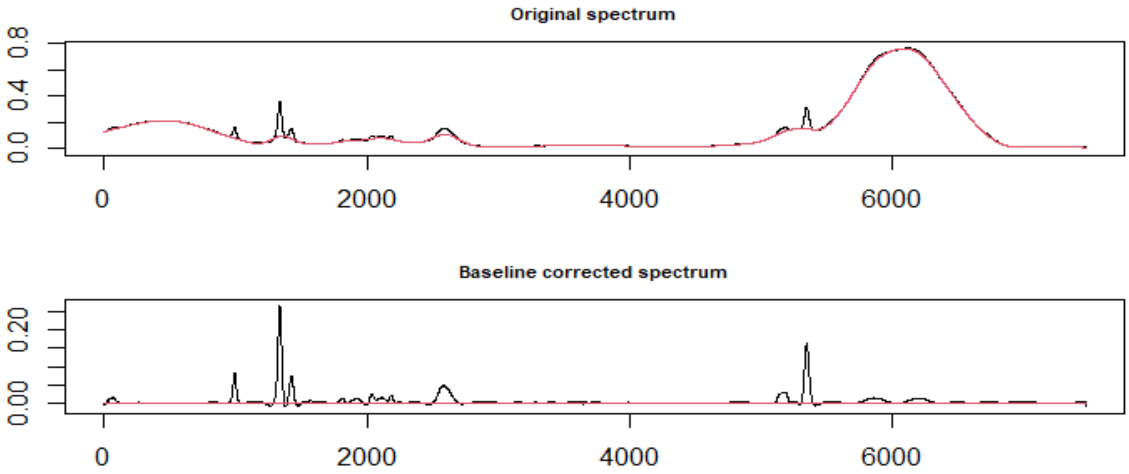


Figura 22: En la figura se muestra el espectro del alcohol artesanal con el que se busca obtener la predicción.

Imágenes elaboradas por el autor.

Así mismo como se realizó con los espectos para la formación del modelo, se realizó la corrección de línea base por medio de los cuadrados asimétricos. Mediante la figura 23 se aprecia una comparativa cuando se aplica la técnica.



Figuras 23: En la figura se muestra el espectro del alcohol artesanal al cual se pasó por un proceso de corrección de línea base vs la comparativa de como se ve antes de esta modificación.

Imágenes elaboradas por el autor.

El paso siguiente fue el suavizado el que se encuentra representado por la figura 24. en el cual se suavizaron los 4 espectros, mediante la metodología establecida previamente.

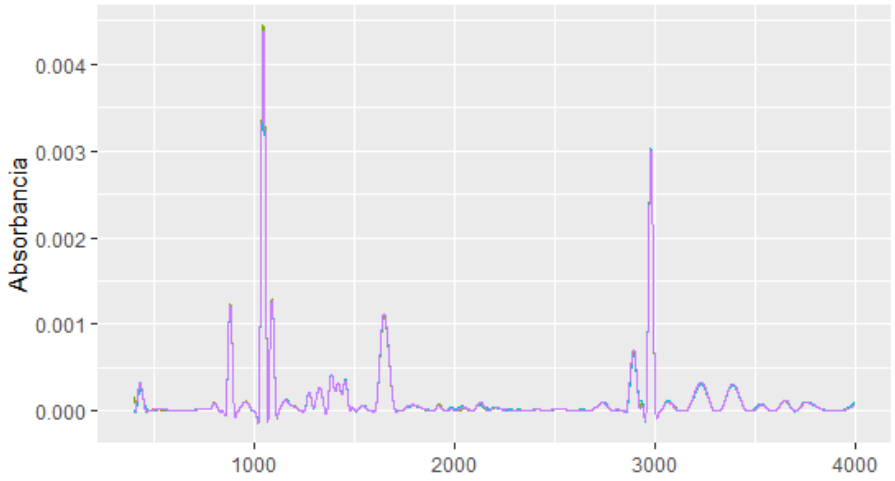


Figura 24: En la figura se muestra el espectro del alcohol artesanal al cual se pasó por un proceso de suavizado.

Imágenes elaboradas por el autor.

Seguido del suavizado procedió a normalizar los espectros como se puede apreciar en la dentro de la figura 25.

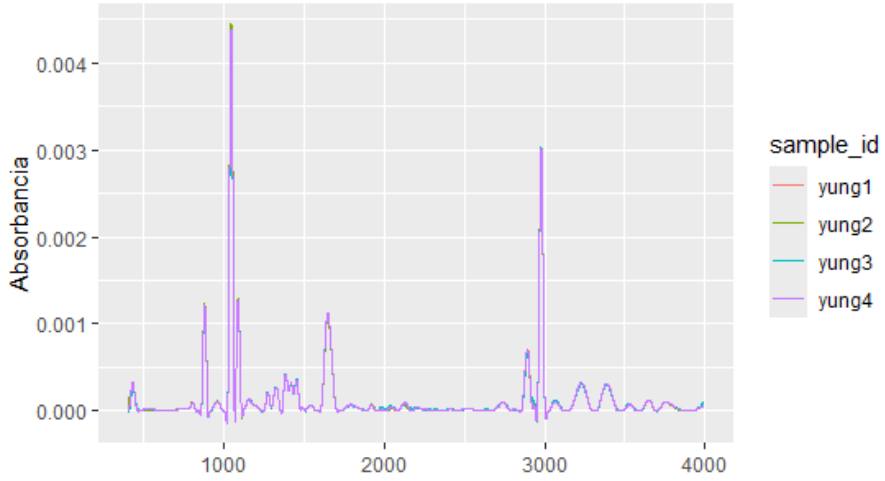
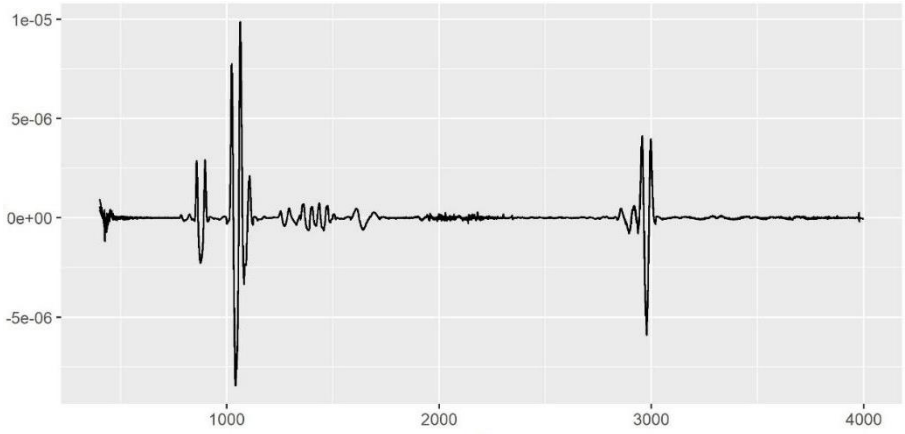


Figura 25: En la figura se muestra el espectro del alcohol artesanal al cual se pasó por un proceso de normalizado.

Imágenes elaboradas por el autor.

y el paso final fue la obtención de la segunda derivada dado que el modelo en que se busca predecir se lo elaboró a partir de ese dato, esta se encuentra retratada dentro de la figura 26.



Figuras 26: En la figura se muestra el espectro del alcohol artesanal al cual se pasó por un proceso para la segunda derivada.

Imágenes elaboradas por el autor.

4.1.3.2 Determinacion de las concentraciones de Etanol-Metanol obtenidas a partir de una muestra de alcohol artesanal

"El laboratorio determinó las concentraciones reales del alcohol mediante un análisis fisicoquímico, cuyos resultados se presentan en la Tabla 2. El documento oficial con los datos completos se encuentra disponible en los anexos al final del informe.

Muestra	Concentración
Etanol	58%
Metanol	0,02 %

Tabla 2: Tabla de Resultados entregados por laboratorio mediante un análisis fisicoquímico.

Imágenes elaboradas por el autor.

Conociendo los valores reales presentes en las muestras y establecido el mejor modelo para la detección simultánea de los alcoholes, el paso final fue agregar los espectros procesados del alcohol artesanal al modelo con el fin de obtener las concentraciones generadas. Estas concentraciones se encuentran presentadas dentro de la tabla 3.

Predicciones obtenidas por modelo de etanol-metanol

Muestra	**Componente 1**		**Componente 2**	
	Etanol	Metanol	Etanol	Metanol
“trago 1”	40.57	0.52	41.31	0.02
“trago 2”	40.63	0.24	40.91	0.04
“trago 3”	40.56	0.56	40.16	0.05
“trago 4”	40.56	0.55	40.82	0.02

Tabla 3: Predicciones elaboradas por el modelo a partir de alcohol artesanal sobre la componente del etanol.

Imágenes elaboradas por el autor.

En la Tabla 1 se presentan los resultados obtenidos de la predicción para cuatro muestras de alcohol artesanal, etiquetadas como “trago 1” a “trago 4” , que en el modelo fueron evaluados dentro del modelo por sus dos componentes esto implica que la primera componente tiene más afinidad hacia la predicción de la variable del etanol mientras, que la segunda componente indica una mayor afinidad hacia la variable del metanol, pero hay que tener en cuenta que las dos componentes predicen de las dos variables.

Se aprecia que la componente 1 en la predicción +del etanol es en promedio 40,58 % y para el metanol una predicción de 0.46 % mientras pero en cambio con la segunda componente la predicción para del etanol fue en promedio 40.8 % peor en cambio para la predicción de metanol fue de 0,03 %.

4.1.3.3 Comparación con el análisis de laboratorio

Comparando las concentraciones obtenidas del laboratorio con los obtenidos a partir del modelo tiene veracidad para la predicción del metanol en la segunda componente lo predecido por el modelo fue de 0.03% en promedio y el valor real dentro de la muestra fue de 0.02% y teniendo en cuenta que el modelo tiene un error mse del 0.380 está acorde al margen de error, mientra que para la detección del etanol este fue infra expresado dado que se la predicción fue en promedio 40.88 mientras que la muestra real tuvo 58 % .

Se puede observar una diferencia entre las concentraciones reales presentes en las muestras y aquellas predichas por el modelo. Al indagar más a fondo sobre una posible fuente de error durante el pretratamiento o en la fase experimental, se concluyó que, al momento de construir el modelo, este fue elaborado utilizando estándares que abarcaban concentraciones de etanol entre el 30 % y el 50 %, y de metanol entre el 0 % y el 2 %. Estos rangos fueron seleccionados con el objetivo de cubrir valores significativamente superiores e inferiores a los establecidos por la norma INEN.

Sin embargo, al incluir muestras con concentraciones fuera de estos rangos, el modelo incurre en un proceso de extrapolación, lo que afecta negativamente su capacidad predictiva, especialmente en la estimación de los niveles de etanol. Dado que el modelo fue diseñado para realizar una detección simultánea de etanol y metanol, la predicción del metanol también se vio comprometida, generando estimaciones inadecuadas.

No obstante, en los casos en que las muestras se encuentran dentro del rango establecido durante el diseño experimental, el modelo demuestra un buen desempeño, con un valor de 0,380 en la métrica utilizada para la predicción simultánea de etanol y metanol.

4.2 Discusión

Dentro del presente estudio se buscó la determinación de pretratamientos en espectros con el fin de la elaboración y el análisis de datos más adecuada para lograr la mejor predicción en la cuantificación simultánea de metanol y etanol, a partir de los espectros obtenidos mediante FTIR con el fin de predecir las concentraciones de alcohol presentes bebidas alcohólicas artesanales producidas en el valle de Yunguilla, para el modelo fue elaborado a través de la técnica de sPLS, el cual mostró un MSE: de 0,360 con 2 componentes lo que asegura una gran capacidad predictiva pero solo establecido para el rango establecido entre 30 % y 50 %, y al ingresar muestra que gracias al análisis de laboratorio se dictaminó que 58 % que no se encuentra dentro del rango de calibración lo que causa una gran pérdida durante la predicción. Por lo que se puede rescatar que él, que lo óptimo es tener el análisis de las concentraciones de alcoholes para la elaboración del rango de calibración o la otra opción es el aumento de los rangos para la elaboración de los estándares lo que generaría, un modelo con mayor capacidad predictiva a costa de mayores recursos computacionales.

Las principales limitaciones dentro del estudio es que las muestras es que dado que no existe un protocolo estricto en la hora de la destilación esto genera que no sea un proceso estandarizado por lo que el volumen de etanol -metanol es posible que sea superior a lo que requiere la normativa lo que generó el error en la hora de la predicción de las concentraciones, por lo que para evitarlo es preciso un mayor rango en la elaboración de los estándares.

4.2.1 Comparación con estudios previos

Alrededor del tema de FTIR existen muchas investigaciones que asimismo como se hizo en este estudio, utilizan el método SPLS para la elaboración de los múltiples modelos con el fin de distinguir distintas variables.

Y un estudio con un enfoque parecido, pero buscando la detección discriminatoria de vinagre y etanol, el estudio elaborado por parte (Gabriela Lisseth García Loja, 2025) señala la elaboración de un modelo formado a través de PLS DA para la detección de presencia de etanol en muestras de vinagre alcanzado una margen del RMSE de 0,11 para la detección del etanol mientras que para el vinagre dio una RMSE de cerca del 0,34 eso indica que el modelo tiene una alta confiabilidad para la predicción. Pero porqué situación no salió en nuestro caso eso se dio porque las concentraciones de ácido acético y etanol eran más pequeñas y estaban en los rangos de detección, a comparación del nuestro el cual se buscó obtener las concentraciones a partir sólo de un rango desconocido.

Un claro ejemplo de esto dentro del estudio aplicable en la industria se del dentro del estudio de(Yaman, 2020), en el cual se empleó espectroscopia de Raman e infrarroja. Para la elaboración de un modelo capaz de detectar muestras de leche vaca que se la hacían pasar como leche de cabra, para elaborarlo lo modelaron a través de través de un modelo de PLS y este alcanzó a una correlación de validación del 0.96.

Otro ejemplo relación al mundo del micro plástico se ve en, Asimismo, en el estudio por parte de (Da Silva et al., 2020) desarrollar un método analítico automatizado que caracterice micro plásticos de un tamaño inferior a 100 micras con el fin de buscar para los plásticos más usados dentro de la industria, en el estudio fue una combinación de herramientas de Deep learning junto a modelos PLS-DA y SIMCA. Siendo el mejor modelo de PLS-DA con una tasa de clasificación de éxito del 95 %. La muestra usada dentro del estudio fue a partir de sedimentos de diferentes mezclas de plástico.

Y uno de los estudios más recientes en el cual se muestra la técnica de PLS y sPLS la cual se usó con el fin de seleccionar variables relevantes en campos que tengan una alta

dimensionalidad con el objetivo de análisis de datos genómicos. Para la optimización de los modelos se empleó un mecanismo en cual está basado en el método LARS en el cual se calculan soluciones de manera rápida. Para la valoración del método se realizaron pruebas de validación cruzada. Los resultados principales mostraron que el mejor modelo al que responden los datos de expresión génica relacionado junto a la unión de factores de transcripción fue mediante el sPLS ya que ayuda a la reducción de las dimensiones en precisión y la capacidad de identificación de variables relevantes.(Chun & Keleş, 2010).

Capítulo 5

5.1 Conclusión

En conclusión, de este estudio se rescatan las principales optimizaciones para la detección simultánea se puede decir que se logró la interacción de los espectros frente a sus distintas concentraciones. Por lo resultados confirman que acorde al procesamiento previo de los espectros generados indica el modelo ayuda al predictivo en la cuantificación del etanol y metanol en bebidas de origen artesanal, entonces podemos rescatar que al usa el espectro completo procesado con la primera derivada tiene una mejor predicción que hacer la predicción de moldeo que si solo se emplea solo los picos de interés. Así se minimiza el error durante la predicción.

Durante el armado del modelo se rescata que tratándose de espectros completos tiene un mejor rendimiento durante para la predicción los modelos formados a través de SPLS que al usar el PLS. Eso se lo evidencia gracias a el RMSE Al integrar el espectro de la bebida de alcohol artesanal al modelo se debe hacer un gran énfasis en que el

volumen alcohol teórico la muestra debe constar dentro de los rangos de los estándares que el modelo requiere para la predicción así se evita la extrapolación del modelo.

5.2 Recomendaciones

Recomienda que dentro de futuras investigaciones se debe realizar un modelo con unos rangos más amplios con el fin de que el modelo sea eficiente frente a muestras con concentración de etanol y metanol más altas por lo que se recomendaría un rango de concentraciones para los estándares de entren los 25 % hasta los 70 % grados de etanol para la detención del metanol no se siguiere cambios dado que está en función a lo que la normal indica. Además, se propone otro tratamiento de los espectros para ver los cambios generados en las cuales se recomienda distintas técnicas como puede ser la Normal variante estándar (SNV)o la corrección multivariable escalar (MSC), y para finalizar, como se ejemplifico dentro del marco teórico se debe escalar la de espectroscopia FTIR a distintas áreas y aplicaciones dentro de la biotecnología gracias a su amplia versatilidad.

Capítulo 6

6.1 Códigos

Código usado y su disponibilidad

Todos los paquetes, funciones, datos espectrales, los rango y script que fueron utilizados dentro de este estudio esta subidos dentro de la plataforma GitHub para su uso en futuras investigaciones.

https://github.com/rickyjaratesis/codigos-tesis.git

Se hace hincapié en revisar el script de nombre final. R

6.2 Anexos

Reporte de laboratorio

MSV LABORATORIO
Análisis de alimentos, aguas y suelos

INFORME DE RESULTADOS

Informe: MSV-IE-1463-25
Orden de ingreso: OI-678-25
Cuenca, 21 de Julio del 2025

DATOS DEL CLIENTE

Cliente: RICARDO JARA
Dirección: EMILIANO ZAPATA Y SN
Teléfono: 0981093972

DATOS DE LA MUESTRA

[2]NOMBRE DE LA MUESTRA: TRAGO DE YUNGUILLA			
[2]MARCA COMERCIAL: NA		**[2]FABRICANTE:** NA	
PROCEDENCIA: YUNGUILLA	**TIPO DE MUESTRA:** BEBIDA ALCOHOLICA	**[2]TIPO DE ENVASE:** PET	
[2]PRESENTACIONES: 450 ml	**[2]FORMA DE CONSERVACION:** AMBIENTE FRESCO Y SECO	**CONDICIONES DE ANALISIS:** TEMPERATURA AMBIENTE T 20.7ºC ±5, HR 54 ±5%	
CODIGO MUESTRA: OI67825	**[2]LOTE:** NA	**[2]FECHA ELAB:** 2025-07-17	**[2]FECHA CAD:**
FECHA RECEPCION: 2025-07-17	**FECHA ANALISIS:** 2025-07-17 - 2025-07-21	**FECHA ENTREGA:** 2025-07-21	
ENSAYO EN: LABORATORIO	**MUESTREO:** CLIENTE	**NUMERO DE MUESTRAS:** UNO (1)	

ENSAYOS ANÁLISIS FISICO-QUIMICOS

PARÁMETRO	MÉTODO - TÉCNICA	UNIDAD	RESULTADO
*GRADO ALCOHOLICO	NTE INEN 340 - DESTILACION	°GL	58
*METANOL	NTE INEN 2014 - CROMATOGRAFIA DE GASES	mg/100cc	<0.02

*Fuera del alcance de la acreditación. **Subcontratado acreditado. ***Subcontratado no acreditado. U:INCERTIDUMBRE.

Dra. Sandra Guaraca
GERENTE DE LABORATORIO

Cualquier información adicional correspondientes a los ensayos que requiera el cliente, están a disposición. Los datos e información de las muestras (tal como se reciben) y de los clientes, que puedan afectar la validez de los resultados han sido proporcionados por el cliente y son de su exclusiva responsabilidad. El Laboratorio no será responsable de los desvíos encontrados en los ítems de ensayo entregados por los clientes que puedan afectar a los resultados, que al ser detectados serán comunicados al cliente.

Los resultados expresados en este informe tienen validez sólo para la muestra recibida en el laboratorio. Este informe no será reproducido sin la aprobación de MSV. [1]Opiniones e interpretaciones están fuera del alcance del SAE. [2]Información proporcionada por el cliente, MSV se responsabiliza exclusivamente de los análisis realizados. Regla de decisión: •CUMPLE: El valor medido está dentro del límite de aceptación, •NO CUMPLE: El valor medido está fuera del límite de aceptación, •NO APLICA: No se tiene requisito de comparación; no se tomará en cuenta la incertidumbre asociada al resultado. MSV está comprometido con la imparcialidad y confidencialidad de la información y los resultados (este informe representa la aceptación de la política declarada de MSV en relación al tema).

Avenida de las Américas y Turuhuaico, Edificio Miraflores, 3er piso. Cuenca-Ecuador
Teléfono: 074045127 Celular: 0995354172 msvlaboratorio@gmail.com

6.3 Bibliografía:

APRENDE ESTADÍSTICAS FÁCILMENTE. (n.d.). *¿Qué es: Tuning? Optimice sus modelos de datos*. Retrieved July 25, 2025, from https://es.statisticseasily.com/glossario/what-is-tuning-optimize-your-data-models/

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

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

Berthomieu, C., & Hienerwadel, R. (2009b). 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

Chun, H., & Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society.*

Series B, Statistical Methodology, *72*(1), 3. https://doi.org/10.1111/J.1467-9868.2009.00723.X

Da Silva, V. H., Murphy, F., Amigo, J. M., Stedmon, C., & Strand, J. (2020). Classification and Quantification of Microplastics (<100 μm) Using a Focal Plane Array-Fourier Transform Infrared Imaging System and Machine Learning. *Analytical Chemistry*, *92*(20), 13724–13733. https://doi.org/10.1021/ACS.ANALCHEM.0C01324/SUPPL_FILE/AC0C01324_SI_004.PDF

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

Fleming Patrick. (n.d.). *5.7: Espectroscopia - Química LibreTexts*. Retrieved July 25, 2025, from https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Quantum_Chemistry_with_Applications_in_Spectroscopy_(Fleming)/05%3A_The_Rigid_Rotor_and_Rotational_Spectroscopy/5.07%3A_Spectroscopy

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.

Granados, R. M. (n.d.). *Modelos de regresión lineal múltiple*.

Hyonho Chun, & Sündüz Keleş. (2010, January). *Sparse partial least squares regression for simultaneous dimension reduction and variable selection - PMC*. https://pmc.ncbi.nlm.nih.gov/articles/PMC2810828/

INEN. (2013). *Norma Técnica Ecuatoriana NTE INEN 2296: Primera Revisión*.

Juan V. Ashurst, David H. Schaffer, & Thomas M. Nappe. (2025, January 6). *Toxicidad del metanol - StatPearls - Estantería NCBI*. https://www.ncbi.nlm.nih.gov/books/NBK482121/

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

Manuel Vázquez Vázquez, D. (n.d.). *UNIVERSIDAD DE SANTIAGO DE COMPOSTELA*.

Ministerio de Salud Pública. (n.d.). *Actualización de casos atendidos por consumo de alcohol metílico – Ministerio de Salud Pública*. Retrieved July 23, 2025, from https://www.salud.gob.ec/actualizacion-de-casos-atendidos-por-consumo-de-alcohol-metilico-3/

Ministerio de Salud Pública. (2020). *La cifra de fallecidos por intoxicación con alcohol metílico asciende a 24 – Ministerio de Salud Pública*. https://www.salud.gob.ec/la-cifra-de-fallecidos-por-intoxicacion-con-alcohol-metilico-asciende-a-24/

Nguyen, D. V, & Rocke, D. M. (2002). Tumor classification by partial least squares using microarray gene expression data. In *BIOINFORMATICS* (Vol. 18, Issue 1).

Pérez-Enciso, M., & Tenenhaus, M. (2003). 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

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

Random Search and Grid Search for Function Optimization - MachineLearningMastery.com. (n.d.). Retrieved July 22, 2025, from https://machinelearningmastery.com/random-search-and-grid-search-for-function-optimization/

Richman Jeffrey. (n.d.). *Data Normalization Explained: Types, Examples, & Methods | Estuary*. Retrieved July 25, 2025, from https://estuary.dev/blog/data-normalization/

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., Góźdź, S., Macek, P., Kalwat, Z., Khalavka, M., Rzad, 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*(1). 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/.

Villanueva Anadón, B., Ferrer Dufol, A., Civeira Murillo, E., Gutiérrez Cia, I., Laguna Castrillo, M., & Cerrada Lamuela, E. (2002). Intoxicación por metanol. *Medicina Intensiva*, *26*(5), 264–266. https://www.medintensiva.org/es-intoxicacion-por-metanol-articulo-13033600

Yaman, H. (2020). A rapid method for detection adulteration in goat milk by using vibrational spectroscopy in combination with chemometric methods. *Journal of Food Science and Technology*, *57*(8), 3091. https://doi.org/10.1007/S13197-020-04342-4