



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
CUAUTITLÁN**

**“Autenticación de Tequilas del CRT por Categoría
mediante la aplicación de Modelos elaborados a partir de
Técnicas Quimiométricas y Espectrofotometría FT-MIR”**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Licenciado en Química Industrial

P R E S E N T A :

YAIR ELI MOLINA RENTERIA
NO. CTA. 31027172-3

DIRECTOR DE TESIS:

DRA. MARÍA GUADALUPE PEREZ CABALLERO

CUAUTITLÁN IZCALLI, ESTADO DE MÉXICO, 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

**FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN**

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, autoriza al alumno:
Yair Elí Molina Rentería

Con número de cuenta: 310271723 a presentar el: Tesis y examen profesional

Autenticación de Tequitas del CRT por Categoría mediante la aplicación de Modelos elaborados a partir de
Técnicas Quimiométricas y Espectrofotometría FT-MIR

Bajo la asesoría del: Dra. María Guadalupe Pérez Caballero

Para obtener el título de: Licenciado en Química Industrial

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA	FECHA
PRESIDENTE	<u>Dra. María Guadalupe Pérez Caballero</u>		<u>2 -11-2021</u>
VOCAL	<u>Dra. Carolina Moreno Ramos</u>		
SECRETARIO	<u>Dr. Benjamín Velasco Bejarano</u>		
1er. SUPLENTE	<u>M. C. Juan Carlos Rueda Jackson</u>		
2do. SUPLENTE	<u>M. C. Martha Angélica Villegas González</u>		

Atentamente notificamos su participación en la revisión y evaluación del trabajo para que en un plazo no mayor a 15 días hábiles emita su VOTO APROBATORIO. Si existieran correcciones u observaciones en el trabajo, contaría con 15 días hábiles más a partir del momento en que el pasante entrega el trabajo corregido, para emitir su VOTO APROBATORIO o RECHAZO, todo esto de acuerdo con el Reglamento para Titulación de la FES Cuautitlán.

ATENTAMENTE

"POR MI RAZA HABLARÁ EL ESPÍRITU"

Cuautitlán Izcalli, Méx. a 29 de octubre de 2021.

I.A. LAURA MARGARITA CORTAZAR FIGUEROA
JEFA DEL DEPARTAMENTO

NOTA: Los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.

En caso de que algún miembro del jurado no pueda asistir al examen profesional deberá dar aviso por anticipado al departamento.

LMCF/lmcf*



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, autoriza al alumno:
Yair Eli Molina Rentería

Con número de cuenta: 310271723 a presentar el: **Tesis y examen profesional**

Autenticación de Tequilas del CRT por Categoría mediante la aplicación de Modelos elaborados a partir de Técnicas Quimiométricas y Espectrofotometría FT-MIR

Bajo la asesoría del: Dra. María Guadalupe Pérez Caballero

Para obtener el título de: Licenciado en Química Industrial

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA	FECHA
PRESIDENTE	<u>Dra. María Guadalupe Pérez Caballero</u>	_____	_____
VOCAL	<u>Dra. Carolina Moreno Ramos</u>		<u>1-11-2021</u>
SECRETARIO	<u>Dr. Benjamin Velasco Bejarano</u>	_____	_____
1er. SUPLENTE	<u>M. C. Juan Carlos Rueda Jackson</u>	_____	_____
2da. SUPLENTE	<u>M. C. Martha Angélica Villegas González</u>	_____	_____

Atentamente notificamos su participación en la revisión y evaluación del trabajo para que en un plazo no mayor a 15 días hábiles emita su VOTO APROBATORIO. Si existieran correcciones u observaciones en el trabajo, contaría con 15 días hábiles más a partir del momento en que el pasante entrega el trabajo corregido, para emitir su VOTO APROBATORIO o RECHAZO, todo esto de acuerdo con el Reglamento para Titulación de la FES Cuautitlán.

ATENTAMENTE

"POR MI RAZA HABLARÁ EL ESPÍRITU"

Cuautitlán Izcalli, Méx. a 29 de octubre de 2021.


LA. LAURA MARGARITA CORTÁZAR FIGUEROA
JEFA DEL DEPARTAMENTO

NOTA: Los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.

En caso de que algún miembro del jurado no pueda asistir al examen profesional deberá dar aviso por anticipado al departamento.

LMCF/lncl*



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

**FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN**

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, autoriza al alumno:
Yair Eli Molina Rentería

Con número de cuenta: 310271723 a presentar el: Tesis y examen profesional

Autenticación de Tequilas del CRT por Categoría mediante la aplicación de Modelos elaborados a partir de Técnicas Quimiométricas y Espectrofotometría FT-MIR

Bajo la asesoría del: Dra. María Guadalupe Pérez Caballero

Para obtener el título de: Licenciado en Química Industrial

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA	FECHA
PRESIDENTE	<u>Dra. María Guadalupe Pérez Caballero</u>	_____	_____
VOCAL	<u>Dra. Carolina Moreno Ramos</u>	_____	_____
SECRETARIO	<u>Dr. Benjamín Velasco Bejarano</u>		<u>29/10/2021</u>
1er. SUPLENTE	<u>M. C. Juan Carlos Rueda Jackson</u>	_____	_____
2do. SUPLENTE	<u>M. C. Martha Angélica Villegas González</u>	_____	_____

Atentamente notificamos su participación en la revisión y evaluación del trabajo para que en un plazo no mayor a 15 días hábiles emita su VOTO APROBATORIO. Si existieran correcciones u observaciones en el trabajo, contaría con 15 días hábiles más a partir del momento en que el pasante entrega el trabajo corregido, para emitir su VOTO APROBATORIO o RECHAZO, todo esto de acuerdo con el Reglamento para Titulación de la FES Cuautitlán.

ATENTAMENTE

"POR MI RAZA HABLARÁ EL ESPÍRITU"

Cuautitlán Izcalli, Méx. a 29 de octubre de 2021.


I.A. LAURA MARGARITA CORTAZAR FIGUEROA
JEFA DEL DEPARTAMENTO

NOTA: Los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.

En caso de que algún miembro del jurado no pueda asistir al examen profesional deberá dar aviso por anticipado al departamento.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

**FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN**

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, autoriza al alumno:
Yair Eli Molina Rentería

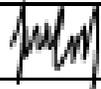
Con número de cuenta: 310271723 a presentar el: **Tesis y examen profesional**

Autenticación de Tequilas del CRT por Categoría mediante la aplicación de Modelos elaborados a partir de Técnicas Quimiométricas y Espectrofotometría FT-MIR

Bajo la asesoría del: Dra. María Guadalupe Pérez Caballero

Para obtener el título de: Licenciado en Química Industrial

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA	FECHA
PRESIDENTE	<u>Dra. María Guadalupe Pérez Caballero</u>	_____	_____
VOCAL	<u>Dra. Carolina Moreno Ramos</u>	_____	_____
SECRETARIO	<u>Dr. Benjamín Velasco Bejarano</u>	_____	_____
1er. SUPLENTE	<u>M. C. Juan Carlos Rueda Jackson</u>		<u>01/11/2021</u>
2do. SUPLENTE	<u>M. C. Martha Angélica Villegas González</u>	_____	_____

Atentamente notificamos su participación en la revisión y evaluación del trabajo para que en un plazo no mayor a 15 días hábiles emita su VOTO APROBATORIO. Si existieran correcciones u observaciones en el trabajo, contaría con 15 días hábiles más a partir del momento en que el pasante entrega el trabajo corregido, para emitir su VOTO APROBATORIO o RECHAZO, todo esto de acuerdo con el Reglamento para Titulación de la FES Cuautitlán.

ATENTAMENTE

"POR MI RAZA HABLARÁ EL ESPÍRITU"

Cuautitlán Izcalli, Méx. a 29 de octubre de 2021.


LA. LAURA MARGARITA CORTAZAR FIGUEROA
JEFA DEL DEPARTAMENTO

NOTA: Los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.

En caso de que algún miembro del jurado no pueda asistir al examen profesional deberá dar aviso por anticipado al departamento.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, autoriza al alumno
Yair Eli Molina Rentería

Con número de cuenta 310271723 a presentar el: **Tesis y examen profesional**

Autenticación de Tequilas del CRT por Categoría mediante la aplicación de Modelos elaborados a partir de
Técnicas Quimiométricas y Espectrofotometría FT-MIR

Bajo la asesoría del: Dra. María Guadalupe Pérez Caballero
Para obtener el título de: Licenciado en Química Industrial

PROFESORES QUE INTEGRAN EL JURADO

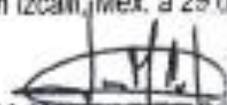
	NOMBRE	FIRMA	FECHA
PRESIDENTE	<u>Dra. María Guadalupe Pérez Caballero</u>	_____	_____
VOCAL	<u>Dra. Carolina Moreno Ramos</u>	_____	_____
SECRETARIO	<u>Dr. Benjamin Velasco Bejarano</u>	_____	_____
1er. SUPLENTE	<u>M. C. Juan Carlos Rueda Jackson</u>	_____	_____
2do. SUPLENTE	<u>M. C. Martha Angélica Villegas González</u>		<u>09/10/2021</u>

Atentamente notificamos su participación en la revisión y evaluación del trabajo para que en un plazo no mayor a 15 días hábiles emita su VOTO APROBATORIO. Si existieran correcciones u observaciones en el trabajo, contaría con 15 días hábiles más a partir del momento en que el pasante entrega el trabajo corregido, para emitir su VOTO APROBATORIO o RECHAZO, todo esto de acuerdo con el Reglamento para Titulación de la FES Cuautitlán.

ATENTAMENTE

"POR MI RAZA HABLARÁ EL ESPÍRITU"

Cuautitlán Izcalli, Méx. a 29 de octubre de 2021.


LA. LAURA MARGARITA CORTAZAR FIGUEROA
JEFA DEL DEPARTAMENTO

NOTA: Los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.

En caso de que algún miembro del jurado no pueda asistir al examen profesional deberá dar aviso por anticipado al departamento.

Agradecimientos y Dedicatorias

En primer lugar, quiero expresar mi agradecimiento a la directora de esta tesis de licenciatura, Dra. María Guadalupe Pérez Caballero, por la dedicación, conocimiento y apoyo que ha brindado en este trabajo, por su consejo en temas académicos, así como personales que me han permitido llegar a este punto. Gracias por la confianza ofrecida desde que comencé haciendo mi servicio social y por compartirme esa pasión por la investigación y la química analítica. Y ya que un trabajo de investigación de este tipo es siempre fruto de ideas, proyectos y esfuerzos previos que corresponden a otras personas; también le agradezco doctora por sus trabajos realizados que son la base para poder lograr esto.

Asimismo, agradezco a mis compañeros de la carrera de Química Industrial, cada momento vivido dentro y fuera de los salones de clases ha aportado un granito de arena a mi formación como químico y como persona.

Por su orientación, conocimientos y experiencias compartidas agradezco de manera general a los académicos de la FES Cuautitlán los cuales en su mayoría están en pro de la formación de profesionistas preparados, competentes y consientes que ayuden a mejorar la situación actual del país.

Finalmente, gracias a la Universidad Nacional Autónoma de México y a la FES Cuautitlán Campus 1 que me han dado todo en el ámbito profesional; desde una formación académica en el ramo de la química, hasta una formación crítica con la cual se nos exhorta a cuestionar las injusticias y luchar por lo que es correcto siempre trabajando con honestidad y valores por el bien del país y la sociedad mexicana en general.

Por otro lado, un trabajo de investigación es también fruto del apoyo vital que nos ofrecen las personas que nos estiman, sin el cual no tendríamos la fuerza, energía y coraje que nos anima a crecer como personas y como profesionales, por esto me gustaría agradecer y reconocer a las siguientes personas:

Gracias a mi familia, a mi padre y a mi madre que durante toda mi vida me han alentado y apoyado en todos sentidos para tener metas y cumplirlas; gracias por enseñarme que todo es alcanzable cuando se trabaja duro y que a veces es mejor dar un paso hacia atrás y sentar buenas bases a seguir adelante cuando todo es incierto. Especialmente, gracias a mi mamá por ser ese ejemplo de superación en la vida, por ser una luchadora incansable que siempre vio por el bienestar de sus hijos, que les inculcó ese amor por el estudio y nos enseñó que para ser una persona de bien el conocimiento es una herramienta importante.

A mis hermanos Alan y Estefania también les agradezco, ya que, si bien en esta familia no somos personas tan sentimentales, siempre encontramos la forma de demostrarnos amor con nuestras acciones. Estefania eres la más grande de los tres y la primera profesionista de la familia, gracias por abrirnos camino. Alan eres el menor, pero tu habilidad e inteligencia siempre me han alentado a dar lo mejor de mí en cada situación pues desde pequeños ha existido esa competencia sana entre los dos que creo yo nos ha permitido colaborar y ser

parte del crecimiento como persona de ambos. Por todo eso y más les agradezco ya que eso ha sido un aliento para seguir escribiendo este trabajo y así culminar esta etapa de mi vida.

Gracias a mis amigos, que siempre me han prestado un gran apoyo moral y humano, necesarios en los momentos difíciles de este trabajo y esta profesión.

Pero, sobre todo, gracias a ti Marlene mi pareja, mi confidente, mi mejor amiga y el amor de mi vida. Cuando pienso en el tiempo juntos, me parece tan corto, pero a tu lado he aprendido lo que a una persona le toma una vida. Tu apoyo incondicional en cada paso que doy, en cada decisión y adversidad son una bendición en mi vida; me haces querer ser una mejor persona, y por ti es que cada decisión que tomo va encaminada a hacer de este mundo algo mejor para ti y las mujeres del mundo; gracias a ti es que estoy llevando a su culminación este trabajo de investigación pues tu amor me inspira y motiva. Gracias por compartir conmigo esa pasión por la ciencia y alimentar ese amor por el conocimiento, pero sobre todo, gracias por ser tú.

A todos, muchas gracias.

Índice

1.	Resumen.....	12
2.	Prólogo.....	12
3.	Introducción.....	13
4.	Objetivos.....	14
4.1	Objetivo general.....	14
4.2	Objetivos particulares.....	14
5.	Marco Teórico.....	15
5.1	El tequila.....	15
5.1.1	Antecedentes y proceso de elaboración.....	15
5.1.2	Clases y categorías de tequila.....	17
5.1.3	Principales compuestos del tequila.....	18
5.1.4	Denominación de Origen.....	19
5.1.5	Consejo Regulador del Tequila (CRT).....	20
5.1.5	20
5.2	Espectroscopia Infrarrojo.....	20
5.2.1	Fundamento de la espectroscopia Infrarrojo.....	21
5.2.2	Vibraciones moleculares.....	22
5.2.3	Medición en el infrarrojo.....	24
5.2.4	Interpretación y asignación de bandas.....	25
5.2.5	Instrumentación en espectroscopia Infrarrojo.....	26
5.2.6	Espectroscopia FT-IR.....	27
5.2.7	Espectrometría de Reflectancia.....	28
5.2.8	Reflectancia total atenuada (ATR).....	29
5.3	Quimiometría.....	29
5.3.1	Principios de quimiometría.....	30
5.3.2	Análisis Clasificadorio y Modelado.....	33
5.3.3	Métodos quimiométricos no supervisados.....	34
5.3.4	Métodos quimiométricos supervisados.....	36
6.	Metodología.....	40
6.1	Actividades.....	40
6.1.1	Caracterización de bandas.....	40
6.1.2	Modelos de clasificación.....	41

	11
7. Experimentación.....	41
7.1 Diseño práctico.	41
7.1.1 Materiales, reactivos y equipos.	41
7.1.2 Muestras de tequila.....	42
7.1.3 Condiciones experimentales.....	42
7.1.4 Softwares.	42
7.2 Obtención de espectros FT-MIR de muestras de tequilas para su comparación y la construcción de modelos.	43
7.2.1 Pretratamiento de datos.....	43
7.2.2 Construcción de matrices.....	45
8. Resultados y Discusión.....	50
8.1 Análisis comparativo e interpretación de los espectros MIR de un tequila 100 % agave y un mixto.....	50
8.2 Modelos exploratorios con técnicas no supervisadas.	54
8.2.1 Análisis de componentes principales (PCA).....	54
8.2.2 Análisis de Loadings	62
8.2.3 Análisis Cluster (CA).....	64
8.3 Modelos obtenidos bajo técnicas supervisadas.....	67
8.3.1 Curvas de Potencia (PC).	67
8.3.2 Funciones de Potencia (PF)	69
8.3.3 Vecino más cercano (KNN).....	73
8.3.4 Analogía de clases por modelación suave e independiente (SIMCA).	74
9. Conclusiones.	85
10. Bibliografía.....	87
10.1 Referencias.....	87
10.2 Índice de imágenes	89
11. Anexos.....	91
11.1 Abreviaturas.....	91
11.2 Definiciones.....	91
11.3 Reactivos	93
11.4 Espectros MIR de componentes del tequila.....	94
11.5 Parámetros obtenidos para cada modelo construido en Matlab.	102

1. Resumen.

Las adulteraciones en las bebidas alcohólicas son un problema creciente en la industria del tequila, por lo cual, organizaciones como el Consejo Regulador del Tequila (CRT) han tenido que enfocar sus esfuerzos en el desarrollo de técnicas de análisis que ayuden a diferenciar estas bebidas adulteradas o “falsas” con una mayor precisión. Bajo este precepto fue desarrollado el siguiente trabajo en colaboración con el CRT, quienes proporcionaron 63 muestras de tequila blanco (35 muestras de tequila 100% agave y 28 de tequila mixto), las cuales fueron analizadas mediante espectrofotometría FTIR y técnicas quimiométricas; construyendo así con los espectros obtenidos, modelos que permitieran detectar diferencias entre las dos categorías (tequila 100% agave y mixto). Para la construcción de los modelos exploratorios se utilizó PCA y CA, técnicas con las que se detectó tequilas anómalas como el TBM03 y TBM34 los cuales sesgan los modelos por su alto y bajo contenido alcohólico por lo que fueron eliminados para las técnicas posteriores; también mediante el análisis de Loadings se detectó diferentes zonas espectrales en las cuales se encuentran compuestos como alcoholes de cadena larga y furfural que contribuyen a la diferenciación de las dos categorías ya que su concentración es mayor en tequilas 100% agave. Estas zonas fueron identificadas entre 400cm^{-1} y 1550cm^{-1} , con los datos de estas zonas se construyeron los modelos con las técnicas de quimiométricas CP, KNN, PF y SIMCA, los cuales arrojaron valores de exactitud, especificidad y precisión mayores a 0.9, lo cual indica que los modelos son capaces de detectar las diferencias entre las muestras de las diferentes categorías. Con esto se concluye que las técnicas quimiométricas empleadas en las condiciones de trabajo establecidas, ayudan a identificar diferencias entre los tequilas 100% agave y mixtos, diferencias que pueden ser aplicadas a la identificación de tequilas “falsos” o adulterados.

2. Prólogo.

El tequila es la bebida nacional por excelencia, con mucha historia y tradición que no sólo nos recuerda nuestras raíces e identidad como nación, sino que también cuenta con una gran aceptación en el extranjero. En la actualidad el tequila abarca gran parte del mercado de las bebidas alcohólicas tanto en México como en el mundo. Esta bebida se produce de la doble destilación del mosto fermentado obtenido del corazón del “agave tequilana Weber variedad azul”, un tipo de agave que por las condiciones de suelo y clima es endémico de México.

En México existen cerca de 200 variedades de agave, cada una con características únicas. De algunas de ellas se obtiene el mezcal que es también una bebida autóctona muy popular. El tequila es un tipo de “mezcal” que recibe su nombre de la población donde nace; el pueblo de Tequila que está ubicado en Guadalajara y en el cual su gente ha mantenido viva la tradición de esta bebida desde mediados del siglo XVI.

Al pasar de los años la industrialización del proceso y la competencia entre las diferentes casas tequileras requirió el nacimiento de una organización encargada de promover la calidad y proteger la cultura y prestigio de la bebida; es así como nace el Consejo Regulador del Tequila por sus siglas CRT y el cual se consagra como la máxima autoridad en cuanto a verificación y certificación de tequila se refiere.

Con base en datos del Consejo Regulador del Tequila (CRT), entre mayo y enero de 2016 la producción de tequila pasó de 96.4 millones de litros a 118.3 millones de litros, tan solo en el

periodo enero–mayo de este año, la exportación de tequila aumentó 8.8% en volumen, comparado con el mismo lapso de 2015, lo que beneficia a 17 mil productores mexicanos y 160 empresas tequileras [17].

El mundo y la industria del tequila están en constante cambio lo que hace necesario el surgimiento de técnicas más eficientes y de bajo costo que nos permitan asegurar y salvaguardar la calidad de nuestra bebida que es parte de la herencia que tiene México para el mundo.

Durante los últimos años se han utilizado diversas técnicas para el análisis de los tequilas y sus características, técnicas como la Cromatografía de Gases, HPLC y espectrometría de masas han sido utilizadas para caracterizar compuestos presentes en el tequila. La quimiometría es una herramienta que en conjunto con las técnicas espectroscópicas ofrece una alternativa nueva e innovadora que pretende ofrecer un apoyo a los métodos y pruebas utilizados por el CRT para la autenticación del tequila. Hoy en día, son más frecuentes las adulteraciones en los tequilas y el surgimiento de bebidas alternativas de baja calidad ponen en riesgo la integridad de la industria y las instancias involucradas.

El presente trabajo explora y cimienta las bases en el uso de técnicas quimiométricas para la categorización de los tequilas, es también una continuación de los trabajos presentados en los congresos de 2017 y 2018 de la Asociación Mexicana de Química Analítica (AMQA); en los cuales se presentaron avances obtenidos en la diferenciación entre tequilas 100% agave y mixtos. En las siguientes páginas se desarrolla en detalle los modelos obtenidos para este fin por diferentes técnicas quimiométricas.

3. Introducción.

La NOM-006-SCFI-2012 define las categorías para el Tequila con base en el porcentaje de los azúcares provenientes del agave utilizados en su elaboración de la manera siguiente [1]: Tequila 100% agave es el tequila elaborado con azúcares exclusivos del Agave Tequilana Weber variedad azul, mientras que el “Tequila mixto” o simplemente “Tequila” es el producto cuyo jugo de fermentación es enriquecido y mezclado con azúcares provenientes de otras fuentes en un porcentaje no mayor al 49 % [1].

Existen varias técnicas analíticas que han sido reportadas para discriminar entre Tequila 100% agave y Tequila mixto. El uso de las relaciones de isótopos estables [2, 3, 4, 5], han conducido a la discriminación de estas dos categorías mediante la evaluación de las relaciones O^{18}/O^{16} y C^{13}/C^{12} de etanol vía el análisis por SPME-HRGC-IRMS [10]. Sin embargo, las técnicas son sofisticadas y sólo están disponibles en un número limitado de laboratorios. En otros estudios, se ha realizado la determinación de las concentraciones de metanol y 2-/3-metil-1-butanol por Cromatografía de Gases y se ha evaluado su correlación para evaluar la autenticidad de muestras de tequila. Asimismo, se ha constatado que el contenido de compuestos volátiles, responsables del aroma al tequila, tales como fenil etanol, α -terpineol, furfural, benzaldehído, se encuentran en un alto contenido en tequilas 100% agave con respecto a los tequilas mixtos, lo cual depende de la materia prima usada en su producción. No obstante, las determinaciones de CVPDB (Cromatografía de gases y espectrometría de masas por relación isotópica de Deuterio) para el análisis de las relaciones isotópicas de C^{13}/C^{12} por la espectrometría de masas (IR MS) así como

las relaciones de Deuterio/Hidrógeno (D/H) de etanol por SNIF-NMR [11], mostraron pocas posibilidades. [14].

En este trabajo fueron estudiadas 63 muestras proporcionadas por el Consejo Regulador del Tequila (CRT) con el fin de diferenciar entre tequilas 100% agave y mixtos. Fueron utilizadas técnicas quimiométricas supervisadas y no supervisadas en conjunto con la espectroscopia FT-MIR. Con el uso de las técnicas no supervisadas como el Análisis de Componentes Principales (PCA) y Análisis Cluster (CA) fueron establecidos los modelos exploratorios que permitieron identificar muestras anómalas y sirvieron como base para los modelos elaborados por técnicas supervisadas. Así mismo las técnicas de análisis multivariante supervisadas utilizadas para la categorización de las muestras fueron: Curvas de Potencia (CP), Vecino más cercano (K-NN), Funciones de Potencia y Análisis de clases por delación suave e independiente (SIMCA).

Los espectros fueron obtenidos en un espectrofotómetro Perkin Elmer Frontier FT-IR con UATR en un intervalo de 400 a 4000 cm^{-1} midiendo un blanco contra aire entre cada muestra, los datos fueron tratados con el programa Spectrum©. Las matrices utilizadas se construyeron partiendo de datos brutos en todo el intervalo. Los modelos óptimos se construyeron utilizando diferentes intervalos de número de onda y pretratamientos. El análisis de loadings permitió identificar compuestos y su influencia en las diferentes zonas espectrales para la diferenciación. Se logró la diferenciación de los tequilas mixtos y 100% agave con los modelos óptimos obtenidos por las diferentes técnicas quimiométricas.

4. Objetivos.

4.1 *Objetivo general.*

Establecer métodos alternativos de análisis químico para garantizar la autenticidad del tequila 100 % agave y mixto de muestras de tequilas del CRT mediante la aplicación de diferentes modelos construidos a través de técnicas quimiométricas supervisadas y no supervisadas a partir de datos espectrofotométricos del Infrarrojo Medio (MIR).

4.2 *Objetivos particulares.*

- Categorizar los tequilas proporcionados por el CRT en 100% agave y mixtos bajo el uso de técnicas exploratorias para la construcción de modelos no supervisados como lo son Análisis de Componentes Principales (PCA) y Análisis Cluster (CA).
- Categorizar los tequilas del CRT bajo el uso de técnicas supervisadas optimizadas, tales como Curvas de Potencia (PC), Funciones de Potencia (PF), Vecino más cercano (KNN) y Analogía de clases por modelación suave e independiente (SIMCA).
- Determinar los principales compuestos y zonas espectrales que dan la diferenciación entre tequilas 100% agave y mixtos, mediante la utilización de análisis de loadings.

5. Marco Teórico.

5.1 El tequila.

El tequila es la bebida alcohólica mexicana más famosa y de renombre en todo el mundo, y cuya popularidad se disparó en los últimos años tanto a nivel nacional y, sobre todo, a nivel internacional (el tequila se ha convertido en el cuarto espíritu más consumido en el mundo después de whiskey, vodka y ron) [5]. El tequila se elabora a partir de agave *Tequilana Weber* variedad azul, que se cultiva en regiones específicas (sección 4.1.4). La NOM-006-SCFI-2012 define al tequila como: “Bebida alcohólica regional obtenida por destilación de mostos, preparados directa y originalmente del material extraído, en las instalaciones de la fábrica de un Productor Autorizado la cual debe estar ubicada en el territorio comprendido en la Declaración.” [1].

Al tequila también le pueden ser añadidos algunos edulcorantes, colorantes, aromatizantes y/o saborizantes permitidos por la Secretaría de Salud, con objeto de proporcionar o intensificar su color, aroma y/o sabor. A este proceso se le conoce como Abocamiento [1]. A continuación, hablaremos de lo que es la bebida y su contexto en el país.

5.1.1 Antecedentes y proceso de elaboración.

La elaboración del tequila es un proceso que puede variar entre las diferentes casas tequileras, que van desde lo rudimentario hasta un proceso más industrializado; lo cierto es que, cualquiera sea el caso, su fabricación es un proceso artesanal que consta de los siguientes pasos (Imagen 1).



Imagen 1. Proceso de elaboración del tequila.

El proceso completo dura entre 7 y 9 años e inicia con la plantación del agave *Tequilana Weber* variedad azul. El tiempo promedio que requiere la planta para alcanzar su madurez es de 7 años, al término del cual, se alcanza la acumulación máxima de los carbohidratos de esta planta (algunas

tequileras como *Casa Sauza* determinan la cantidad total de azúcares disponibles para la fermentación) [16].

Una vez que la planta de agave alcanzó la madurez, se separan las hojas o pencas de agave del resto de la planta (Jima), estas no son aprovechables por lo que no son necesarias para el proceso de elaboración. El resultado de esta operación es la obtención de la “piña” que es donde se encuentran almacenados los carbohidratos utilizados para la elaboración del Tequila [16].



Imagen 2. Agricultores y jimadores de agave.

La plantación y la jima son los primeros pasos en la elaboración del tequila y ya sea que la tequilera cuente con sus propios campos agrícolas o recurra a agricultores externos, el proceso hasta este punto es el mismo para ambos casos.

La extracción e hidrólisis (cocimiento) son procesos que continúan después de la obtención de la materia prima (la extracción puede ser previa o posterior a la hidrólisis esto dependerá del proceso específico de cada Casa Tequilera).

La hidrólisis es un proceso por el cual se ponen disponibles para la fermentación los azúcares de la piña; se lleva a cabo en hornos de mampostería o autoclaves utilizando vapor a una temperatura entre 100 y 110°C. Con base en la NOM-006-SCFI-2012 se define la hidrólisis como: “un proceso químico, térmico, enzimático o la combinación de los anteriores para obtener azúcares simples aptos para su fermentación” [1].



Imagen 3. Fibra de Agave

En la extracción se separan los carbohidratos contenidos en las fibras de la piña; este paso se realiza comúnmente con una combinación de equipos como desgarradores y molinos de rodillo.

El resultado de la extracción es un jugo al cual con la adición de levaduras y un ajuste de pH y temperatura se le llama “Mosto Fresco”.

La fermentación es el proceso por el cual se transforman los mostos, por acción de las levaduras se obtiene alcohol etílico y dióxido de carbono. En esta etapa, también se formarán otros compuestos que contribuirán a las características sensoriales finales del tequila [16].

La fermentación se lleva a cabo según la siguiente ecuación:



Los factores críticos que se controlan en esta fase son: la temperatura (32 a 35°C), el pH (4-5), nutrientes y la contaminación por organismos que representan una competencia para las levaduras [16].

El tequila es la única bebida alcohólica que pasa por dos procesos de destilación. En la primera destilación se separan los principales constituyentes del mosto, durante la cual se elimina agua y componentes no deseables como son los restos de levaduras, las sales nutritivas, sólidos, algunos alcoholes secundarios como el metanol y un grupo de compuestos conocidos como alcoholes superiores. La destilación se lleva a cabo en alambiques o también comúnmente en columnas de destilación [16]. La segunda destilación es una rectificación donde se obtiene finalmente el tequila. Algunas Casas Tequileras como proceso adicional después de la destilación filtran su producto con un filtro de carbón que remueve el exceso de compuestos grasos que pueden dar una apariencia turbia al tequila. Otros procesos de filtrado aseguran que no haya partículas presentes en el líquido y dan el acabado brillante a nuestros tequilas [15].

Por último, el añejamiento o maduración es un proceso que se lleva con base en la norma [16] en barricas de roble o encino bajo condiciones de temperatura y humedad controlada.

5.1.2 Clases y categorías de tequila.

La NOM-006-SCFI-2012 diferencia a los tequilas con base en su tiempo de añejamiento (clase) y el contenido de azúcares provenientes del agave (categoría) [1].

Dentro de las clases podemos diferenciar a los tequilas como:

- 1) Tequila Blanco o Plata: Este tequila se obtiene posterior a la segunda destilación, su contenido alcohólico se ajusta con agua de dilución y es embotellado. No es sometido a procesos de maduración en barrica.
- 2) Tequila Joven u Oro: Tequila recién destilado que es susceptible de ser abocado (9.2). También puede resultar de la combinación de un tequila Blanco con Reposado o Blanco-Añejo/Extra Añejo.
- 3) Tequila Reposado: Sujeto a un proceso de maduración de por lo menos dos meses en contacto directo con la madera (roble o encino). De igual forma se considera como tequila Reposado a la mezcla de tequila Reposado/Añejo o Reposado/Extra añejo.
- 4) Tequila Añejo: Su tiempo de maduración es desde un año hasta máximo tres, en contacto directo con barrica de madera de roble o encino.
- 5) Tequila Extra Añejo: Su reposo es mayor de tres años en recipientes de roble o encino.

De acuerdo con el porcentaje de los azúcares provenientes del Agave que se utilice en la elaboración del Tequila, éste se puede clasificar en una de las siguientes categorías [1]:

- 1) Tequila 100% agave: Es el tequila elaborado con azúcares exclusivos del Agave *Tequilana Weber* variedad azul.

- 2) Tequila mixto o Tequila: Producto cuyo jugo de fermentación (mostos) son susceptibles de ser enriquecidos y mezclados conjuntamente en la formulación con otros azúcares provenientes de fuentes alternas en un porcentaje no mayor al 49%.

5.1.3 Principales compuestos del tequila.

Durante la hidrólisis de la piña se obtienen diferentes compuestos, dentro de ellos el principal carbohidrato que contiene el agave es la inulina [16, 18], que posteriormente es convertida en fructuosa para su fermentación [18]. Además de la hidrólisis de la inulina, durante el proceso de cocción del agave se producen muchos compuestos volátiles de Maillard (9.2), principalmente el 5-hidroximetil Furfural [19].

Otros estudios han determinado que el tequila contiene compuestos como el fenil etanol, α -terpineol y benzaldehído los cuales son responsables del aroma [16]. La NOM-006 establece de manera general las cantidades que debe contener el tequila de los principales compuestos identificados los cuales se presentan en la tabla 1 [1].

Tabla 1. Especificaciones Físicoquímicas del Tequila.

Parámetros	Tequila Blanco		Tequila Joven u Oro		Tequila Reposado		Tequila Añejo		Tequila Extra añejo		Método de Ensayo (Prueba) (1)
	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	
Contenido Alcohólico a 293 K (20°C) (%Alc. Vol.)	35	55	35	55	35	55	35	55	35	55	NMX-V-013-NORMEX
Extracto Seco (g/l)	0	0,30	0	5	0	5	0	5	0	5	NMX-V-017-NORMEX
Valores expresados en mg/100 ml de Alcohol Anhidro											
Alcoholes Superiores (alcoholes de peso molecular superior al alcohol etílico o aceite de fusel) (como Alcohol Amílico)	20	500	20	500	20	500	20	500	20	500	NMX-V-005-NORMEX
Metanol (2)	30	300	30	300	30	300	30	300	30	300	NMX-V-005-NORMEX
Aldehídos (como acetaldehído)	0	40	0	40	0	40	0	40	0	40	NMX-V-005-NORMEX
Esteres (como acetato de etilo)	2	200	2	200	2	250	2	250	2	250	NMX-V-005-NORMEX
Furfural	0	4	0	4	0	4	0	4	0	4	NMX-V-004-NORMEX

Así mismo, Norma A. Mancilla en su trabajo *Generation of Maillard Compounds from Inulin during the Thermal Processing of Agave Tequilana Weber Var. Azul* [18] expone con más detalle compuestos producidos durante diferentes etapas de la cocción de la piña.

5.1.4 Denominación de Origen.

La declaración general de la denominación de origen “Tequila” fue publicada en el Diario Oficial de la Federación el 9 de diciembre de 1974 [15]. Y aunque ha tenido modificaciones y adiciones, establece la titularidad del tequila al Estado Mexicano.

Con base en el artículo 156 de la Ley de Propiedad Industrial, se establece como Denominación de Origen a la región geográfica del país en la cual se designa un producto originario de la misma y cuyas características y calidad se deben o son obtenidas exclusivamente del medio geográfico de la misma [15, 16].

Con base en lo anterior, esta definición exige [16]:

- La existencia de una zona geográfica delimitada (propuesta por interesados)
- Un reconocimiento o renombre comprobado del lugar geográfico (declaratoria)
- Condiciones precisas de producción (Norma)
- Tipicidad.



Imagen 4. Plantíos de Agave.

La NOM-006-SCFI-2012 también hace referencia a la denominación de origen y establece al agave *Tequilana Weber* variedad azul como la única especie admitida para la elaboración del tequila cultivado dentro del territorio comprendido en la declaración [1]. Actualmente este territorio está determinado por: El estado de Jalisco (todo su territorio), 30 municipios de Michoacán, 8 municipios de Nayarit, 11 municipios de Tamaulipas y 7 municipios de Guanajuato (imagen 5) [1, 15, 16].

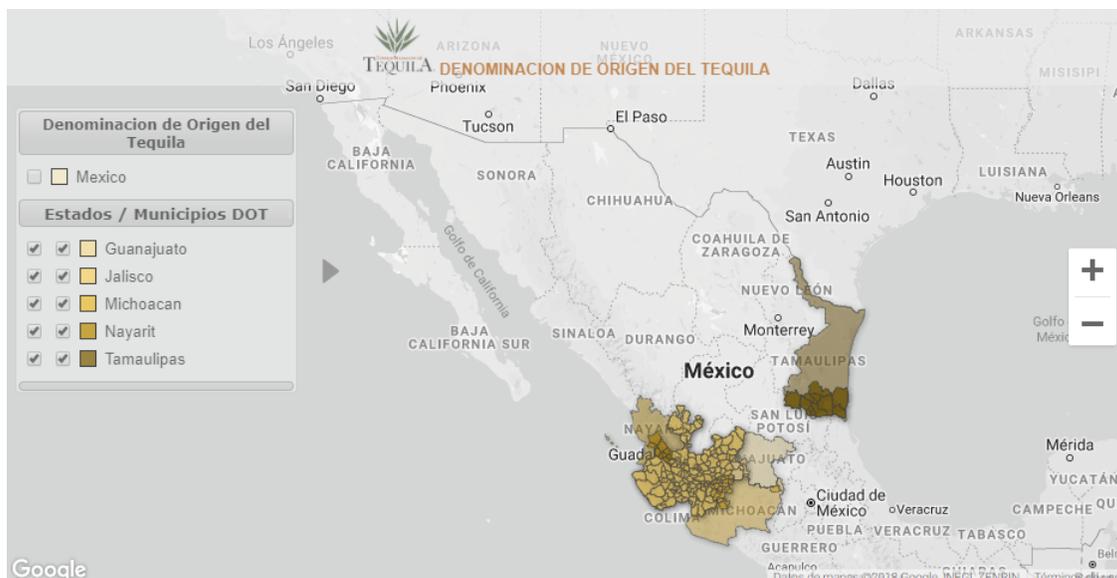


Imagen 5. Mapa de la República Mexicana, Estados y Municipios con DOT marcados en color.

5.1.5 Consejo Regulador del Tequila (CRT).

El Consejo Regulador del Tequila (CRT) es la institución encargada y dedicada a salvaguardar al tequila y todos los actores y agentes productivos ligados a su elaboración. También se encarga de verificar y certificar el cumplimiento de la Norma Oficial del tequila, así como de su calidad [16].

El CRT cuenta con distintas autorizaciones con las cuales asegura la calidad del tequila. Dentro de estas podemos destacar sus certificaciones como: Unidad Verificadora, Organismo de certificación y Laboratorio de pruebas otorgados por la EMA, DGN y COFEPRIS (revisar sección 9.1 para más información) [16].

5.2 Espectroscopia Infrarrojo.

La luz infrarroja es una porción del espectro electromagnético que está justo después de la porción visible de la luz del sol. Esta región puede ser dividida en zonas denominadas infrarrojo cercano “NIR” (Near Infrared), Medio “MIR” (Mid Infrared) y lejano “FIR” (Far Infrared). Cada una de las regiones tiene aplicaciones específicas, pero y en cada una de ellas es el fundamento de la absorción molecular de la luz lo que permite la caracterización [26].

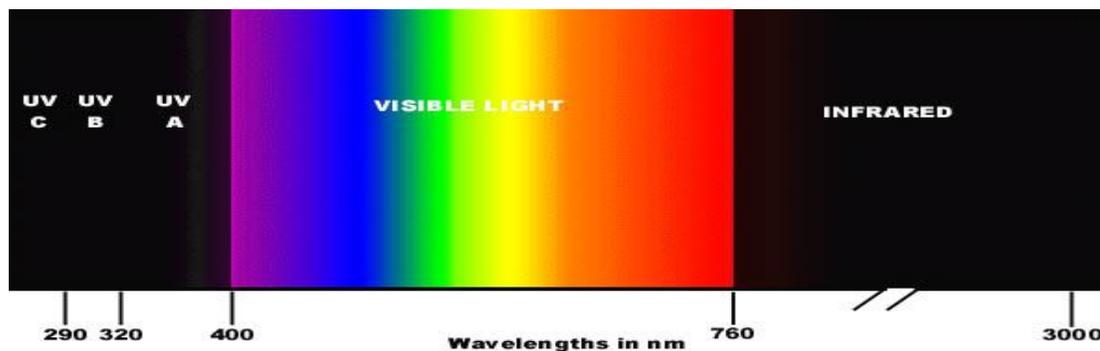


Imagen 6. Zonas del espectro electromagnético.

Esta zona espectral abarca la radiación con números de onda comprendidos entre 12800 y 10 cm^{-1} [26] Las zonas correspondientes a cada región se muestran en la tabla 2.

Tabla 2. Delimitación de zonas espectrales en espectroscopia Infrarrojo.

Sub-región	Número onda (cm^{-1})	Transición característica
Infrarrojo Cercano	12800 - 4000	Sobretonos y combinaciones
Infrarrojo Medio	4000 - 200	Vibraciones fundamentales
Infrarrojo Lejano	200 - 10	Rotaciones

Como se observa en la tabla 2, a cada zona del espectro le corresponde una transición característica que es la responsable de las señales específicas que observamos para cada molécula y cada grupo funcional. Estas transiciones serán explicadas en las siguientes secciones.

La forma en la que se obtiene la información de una muestra típica medida por esta técnica es a través de un gráfico (espectro) que muestra valores de porcentaje de transmitancia (% T) en función del número de onda en unidades de cm^{-1} , esto debido a la proporcionalidad directa que existe entre esta magnitud y la energía o la frecuencia [26].

5.2.1 Fundamento de la espectroscopia Infrarrojo.

En el infrarrojo se pueden obtener espectros de absorción, emisión y reflexión de las distintas especies moleculares que al ser sometidas a un haz de luz de longitud de onda específica producen distintos cambios energéticos en las moléculas (transiciones de estados vibracionales y rotacionales). La absorción de radiación en el infrarrojo es selectiva y se limita en gran parte a especies moleculares para las cuales existen pequeñas diferencias de energía entre los distintos estados vibracionales y rotacionales [26]. En esta zona la radiación no es lo suficientemente energética para poder producir transiciones electrónicas como la radiación ultravioleta y los rayos X.

Las moléculas que absorben en el infrarrojo sufren un cambio neto en el momento dipolar como consecuencia de su movimiento de vibración y rotación; cuando esto sucede, el campo eléctrico alterno puede interactuar con la molécula y de esta manera provocar un cambio en la amplitud de alguno de sus movimientos [26].

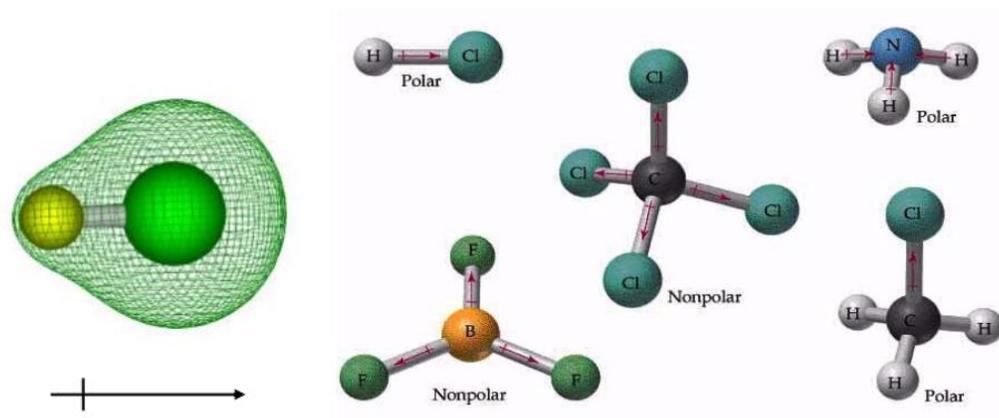
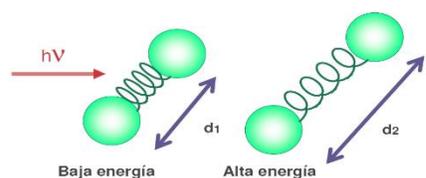


Imagen 7. Momento dipolar, ejemplos de moléculas polares y no polares.

La imagen 7 observamos distintas moléculas con diferente distribución de carga, para el caso del HCl, NH_4^+ y CH_3Cl la distribución es asimétrica, ya que los átomos de cloro y nitrógeno poseen mayor densidad electrónica lo que genera un dipolo y hace a estas moléculas polares. Por el contrario, en las moléculas de BF_3 y CCl_4 la suma vectorial de todos los momentos dipolares se cancela por lo que estas moléculas no son polares y no podrían ser observadas en un espectro infrarrojo. Esto ocurre también con especies mononucleares (O_2 , N_2 o Cl_2), en la cuales el momento dipolar no sufre cambio durante la vibración o rotación.

Siendo así, el momento dipolar está determinado por la magnitud de la diferencia de carga y por la distancia en los centros de carga [26]. En el ejemplo anterior las moléculas de ácido clorhídrico, amonio y cloruro de metilo vibran, y producen una constante variación del momento dipolar, esto origina un campo que puede interactuar con el campo electrónico asociado a la radiación infrarrojo.



Si la frecuencia a la cual vibra la molécula naturalmente es igualada por la frecuencia de la radiación incidente se origina una transferencia de energía que provoca un cambio de amplitud de la vibración molecular [26].

Imagen 8. Cambio de energía molecular

Las transiciones en el infrarrojo se pueden dividir principalmente en dos: Rotacionales (que se dan a niveles de baja energía en el infrarrojo lejano a 100 cm^{-1} o menos) y los vibracionales/rotacionales que se observan en el infrarrojo medio (MIR) [26].

5.2.2 Vibraciones moleculares.

Los átomos de las moléculas se encuentran de manera natural en constante movimiento, esto como consecuencia de la gran cantidad de diferentes tipos de vibraciones y rotaciones alrededor de los enlaces moleculares. En las moléculas sencillas (diatómicas y triatómicas) se puede definir de manera sencilla sus vibraciones en los enlaces y así poder relacionarlas con sus energías de absorción. Cuando hablamos de moléculas poliatómicas el análisis se complica por las interacciones de los centros de vibración [26].

Existen dos diferentes categorías en las que podemos dividir las vibraciones [26]:

- **Tensión o Stretch:** Es un cambio continuo en la distancia interatómica a lo largo del eje del enlace entre dos átomos. En este tipo de vibración podemos tener vibraciones *simétricas* y *anti-simétricas* (imagen 9).



Imagen 9. Tipos de vibraciones de tensión.

- **Flexión o Bend:** Se caracterizan por cambios en el ángulo entre dos enlaces. Dentro de estas encontramos de cuatro tipos, que son de *tijereteo*, *balanceo* u *oscilación*, *aleteo* y de *torsión*.

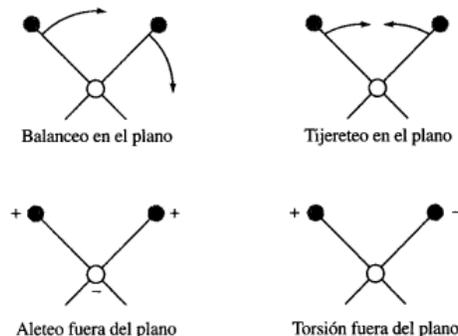


Imagen 10. Tipos de vibraciones de flexión. Nota: el signo + indica movimiento del plano hacia el lector; el signo - indica movimiento del plano alejándose del lector.

En una molécula que tiene más de dos átomos son posibles todos los tipos de vibraciones, además de que estas pueden acoplarse debido a la interacción de los enlaces con un mismo átomo central.

Para poder calcular el número de vibraciones que puede tener una molécula es necesario distinguir entre dos modelos uno que servirá para moléculas sencillas (di-atómicas y triatómicas) y otro para moléculas más complejas o poli-atómicas [26]. El número de vibraciones en moléculas poli-atómicas se calcula considerando que un átomo puede ser localizado utilizando tres coordenadas (x, y, z) por lo tanto, n átomos pueden ser ubicados con 3n coordenadas (3 por cada átomo) (Silverstein et al., 2005). Cada coordenada se comporta como un grado de libertad del átomo, así si una molécula tiene n átomos ella posee 3N grados de libertad (Skoog et al., 2010).

Entonces, el movimiento de una molécula se deriva de los siguientes puntos: (1) del movimiento de traslación (movimiento de la molécula en torno al espacio), (2) movimiento de rotación alrededor de su centro de gravedad y (3) del movimiento de cada uno de los átomos con respecto a los otros (vibraciones) (Skoog et al., 2010; Silverstein et al., 2005). Para poder definir el movimiento de traslación y el de rotación se necesitan tres grados de libertad para cada uno (tres coordenadas para cada caso), el resto de los grados de libertad serán los grados correspondientes

al movimiento de vibración de los átomos, por lo tanto, la expresión adecuada para determinar el número de vibraciones de moléculas poli-atómicas es:

$$3N - 6$$

La ecuación anterior es aplicable para moléculas no lineales, ya que para moléculas lineales todos los átomos se encuentran describiendo una línea recta o sobre el plano, lo que hace imposible la rotación alrededor del eje del enlace y por lo tanto sólo dos grados de libertad describen el movimiento de rotación (Skoog et al., 2010; Silverstein et al., 2005), en ese caso la expresión usada para determinar el número de vibraciones es:

$$3N - 5$$

Como ejemplo tomemos la molécula del agua, la cual consta de 3 átomos ($n=3$) y es una molécula no lineal, por lo tanto, tenemos:

$$3(3) - 6 = 3 \text{ tipos de vibración}$$

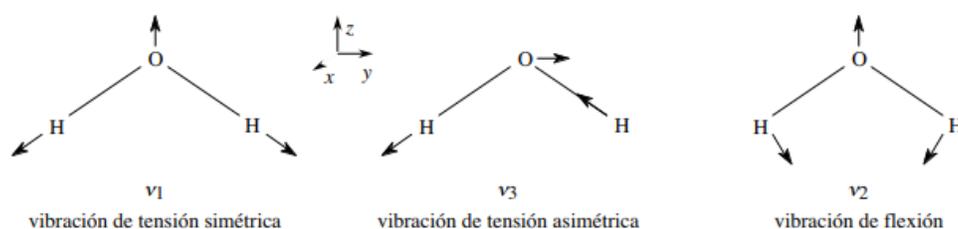


Imagen 11. Vibraciones moleculares del agua.

Las vibraciones de tensión también pueden ser representadas de manera individual por un modelo mecánico sencillo denominado *oscilador armónico*. Este modelo da una buena aproximación ya que consta de dos masas unidas mediante un resorte; si en una de estas masas ocurre una alteración a lo largo del resorte se produce una vibración a la que llamamos *movimiento armónico simple* (imagen 12) [26].

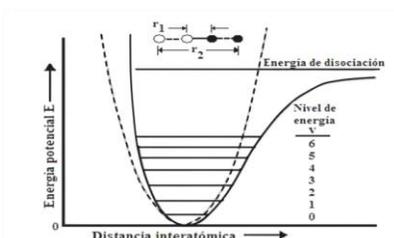


Imagen 12. Diagrama de energía potencial para un oscilador armónico (línea punteada) y un oscilador no armónico (línea continua).

5.2.3 Medición en el infrarrojo.

Al incidir radiación sobre la muestra ésta sufre diferentes fenómenos que dependen de la muestra y la longitud de onda de la radiación. Estos fenómenos son: absorción, transmisión y reflexión. La imagen 13 esquematiza dichos fenómenos donde la luz transmitida por una muestra (P_T), es

menor en intensidad que el haz incidente (P_0); una parte de la intensidad se refleja (P_R) y otra parte es absorbida por la muestra (P_A) [36].

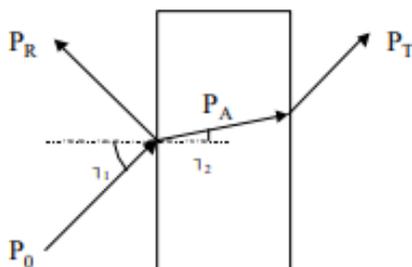


Imagen 13. Representación gráfica de un haz incidente sobre un objeto.

5.2.4 Interpretación y asignación de bandas.

Como se ha mencionado anteriormente el espectro infrarrojo está dividido en tres regiones, de las cuales es el infrarrojo medio la zona en la cual tenemos las frecuencias de grupo. Las bandas observadas en esta zona se pueden asociar a un grupo funcional y a una estructura molecular completa, aunque se debe de considerar que existen estructuras que provocan desplazamientos de las frecuencias de vibración [26, 36].

Las principales vibraciones encontradas en esta zona corresponden que contienen hidrógeno y/o grupos con dobles y triples enlaces aislados. Por otro lado, la asignación de grupos funcionales en la zona de huella digital es más complicada debido a la multiplicidad de bandas [36].

La figura 14 muestra las frecuencias de absorción de los grupos funcionales más comunes en IR medio.

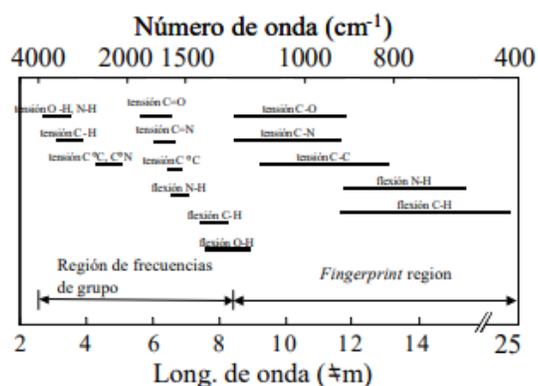


Imagen 14. Frecuencias de vibración en infrarrojo medio.

5.2.5 Instrumentación en espectroscopia Infrarrojo.

Dentro de la espectroscopia de infrarrojo se cuenta con diversos equipos; éstos cuentan con diferentes aditamentos, fuentes de radiación, detectores, dispersores y modelos matemáticos que permiten al analista contar con un equipo específico para sus necesidades. Para poder realizar medidas de absorción en el infrarrojo existen dos instrumentos diferentes [26]:

Tabla 3. Espectrofotómetros utilizados en Infrarrojo.

Equipo	Características	Tipo de espectro	Uso
Dispersivo	Instrumento de doble haz con redes de reflexión para la dispersión de la radiación	Absorción	Análisis cualitativo
No dispersivo	Cuenta con filtro de cuña en vez de dispersante; algunos no cuentan con selector de longitud de onda	Absorción, emisión y reflectancia	Análisis cuantitativo

En la imagen 15 observamos la composición de un espectrofotómetro dispersivo, una fuente de radiación emite energía, ésta pasa por una rejilla (1) que descompone la luz, la radiación se divide en dos haces (2) y el divisor periódico “*Chopping mirror*” (3) deja pasar los dos haces alternadamente. El monocromador (4) cumple la función de filtrar la radiación IR de los haces dispersados; cuando la radiación llega a la muestra (5) se produce una señal eléctrica proporcional a la absorción en el IR, esta señal producida por la muestra llega a un termopar (6) que actúa como detector.

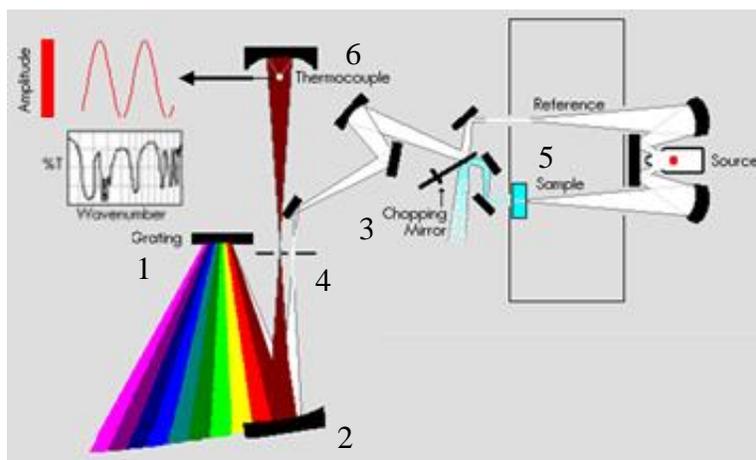


Imagen 15. Esquema de un espectro dispersivo.

Los espectrofotómetros dispersivos son cada vez menos frecuentes en el mercado, en parte por el surgimiento de equipos del tipo multiplex (equipos con transformada de Fourier) que ofrecen una mejor relación calidad-precio y mejor calidad de las señales. Pese a esto, en los equipos dispersivos contamos con diferentes detectores y fuentes de radiación, las más comunes que pueden ser utilizadas comercialmente, se muestran en la tabla 4.

Tabla 4. Aditamentos de un espectrofotómetro dispersivo.

Aditamento	Característica	
Fuente de radiación	Emisor Nernst	Sólido constituido por óxidos de tierras raras alcanza una temperatura de 1500°K, su energía es de 1-10 μm .
	Fuente Globar	Varilla de carburo de silicio. Tiene energías similares al emisor Nernst con temperatura de 1500°K y energía 1-10 μm
	Fuente de filamento	Espiral de alambre de nicromo o Rodio.
	Arco de mercurio	Tubo de cuarzo que contiene un vapor de Hg para IR Lejano ($\lambda > 50 \mu\text{m}$).
	Lámpara de filamento de Tungsteno	Fuente adecuada para IR cercano (4000 a 12800 cm^{-1}) tiene una temperatura de 1100°K y energía de 0.78-2.5 μm .
	Fuente Láser de CO ₂	Produce una banda de absorción en el intervalo 900 a 1100 cm^{-1}
Detectores	Térmico	Depende del poder calorífico de la radiación, se emplean para la detección de todas las longitudes de onda del IR, excepto las más cortas.
	Fotoconductores	Material semiconductor (PbS) la disminución de su resistencia eléctrica al incidir la radiación sirve como medida del haz.
	Piroeléctrico	Láminas de materiales como sulfato de triglicina. Las sustancias piroeléctricas actúan como un condensador.

5.2.6 Espectroscopia FT-IR.

Pertenece al tipo de espectros no dispersivos multiplex y es por mucho el equipo más utilizado en la industria, investigación y docencia ya que representa una gran ventaja en velocidad y exactitud en la obtención de espectros de calidad. Su tecnología se basa en un modelo matemático desarrollado originalmente por Jean-Baptiste Fourier por los años 1800. Por tanto, FT-IR significa Espectroscopia de infrarrojo por transformadas de Fourier (*Fourier Transform Infrared Spectroscopy*) [26, 28].

Los equipos de infrarrojo que trabajan con transformada de Fourier se basan en el interferómetro de Michelson (Imagen 15) [26], que cumple la función de la rejilla de difracción (monocromador) de los espectrofotómetros dispersivos. El interferómetro de Michelson puede dividirse en dos partes: 1.- el interferómetro y 2.- el separador de haz o *beam splitter*. El interferómetro cuenta con un mecanismo de tracción que consiste en dos espejos, uno fijo y el otro móvil que mantiene una velocidad constante respecto a la trayectoria del haz; estos aditamentos son los que permiten obtener interferogramas satisfactorios y en consecuencia espectros de calidad.

Un interferograma (imagen 16) es un gráfico de luz modulada, en el cual tenemos intensidad IR vs posición del espejo móvil o visto de otra forma: señal del detector vs diferencia de paso óptico.

La luz modulada es decodificada usando la transformada de Fourier, obteniendo así el espectro mediante la separación de frecuencias de absorbancia individual creando un gráfico de Intensidad vs Número de onda [28].

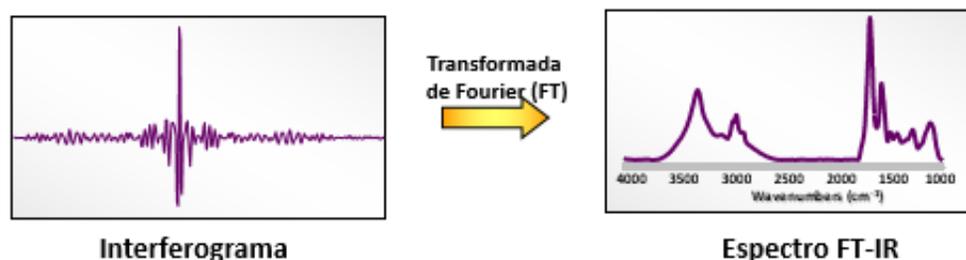


Imagen 16. Interferograma vs Espectro FT-IR.

En la imagen 17 tenemos las principales partes de un espectrofotómetro FT-IR que son: la fuente de radiación, el sistema óptico de dispersión (interferómetro de Michelson), el sistema de detección y registro, portamuestras y algunos otros accesorios como el ATR y el DR [28].

En un espectro por transformadas de Fourier al igual que en dispersivo, la radiación se genera de la fuente (1) y se dirige a través de espejos al separador de haces (2) del interferómetro (3). El rayo láser de He-Ne (4) sigue la radiación IR en su recorrido al interferómetro; al salir del interferómetro, la radiación es dirigida a la muestra (5) y de ahí pasa al detector (6).

De las principales ventajas de los espectrómetros con transformadas de Fourier es el aumento de la relación señal ruido, y la rapidez; esto debido a que la muestra es procesada en pocos segundos (tiempo que tarda en desplazarse el espejo móvil del interferómetro), una alta resolución ($<0.1 \text{ cm}^{-1}$) y la elevada exactitud y reproducibilidad en la determinación de frecuencias. Además, la óptica de este tipo de instrumento permite que todas las radiaciones de la fuente lleguen directo al detector lo que se traduce en una mayor detección de señales inclusive las más débiles [26, 28].

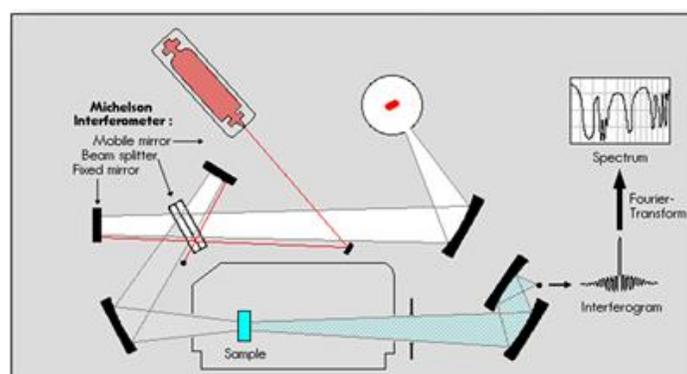


Imagen 17. Esquema de un espectro por transformadas de Fourier.

5.2.7 Espectrometría de Reflectancia.

Cuando un haz de radiación pasa desde un medio más denso a uno menos denso se produce un fenómeno óptico de reflexión, este fenómeno se ve afectado por el ángulo de incidencia del haz

y el sentido de la variación [31]. Hablamos de una reflexión total o especular cuando toda la energía incidente es reflejada en una dirección simétrica al ángulo de incidencia; mientras que la reflexión difusa existe después de que el haz penetra en parte del segundo medio [31], si el medio absorbe la radiación se produce una atenuación del haz en las longitudes de onda de las bandas de absorción lo que se llama reflexión atenuada [26].

La composición del haz reflejado dependerá del índice de refracción del material y de la absorción del compuesto de estudio a determinada longitud de onda.

Los equipos FT-IR cuentan con distintos accesorios de Reflexión especular, reflexión difusa y reflexión total atenuada que se basan en este principio y permiten analizar muestras muy variadas [31]. Dentro de estos aditamentos el más utilizado es el de Reflectancia total atenuada el cual se describe en la siguiente sección.

5.2.8 Reflectancia total atenuada (ATR).

Uno de los aditamentos más utilizado para este tipo de equipos es el ATR (*Attenuated Total Reflectance*). Este es un método de muestreo utilizado para sólidos y líquidos. Su funcionamiento es mediante reflectancia ortogonal por desvanecimiento de onda, aprovecha la diferencia de índices de refracción de la muestra y el cristal del accesorio [28].

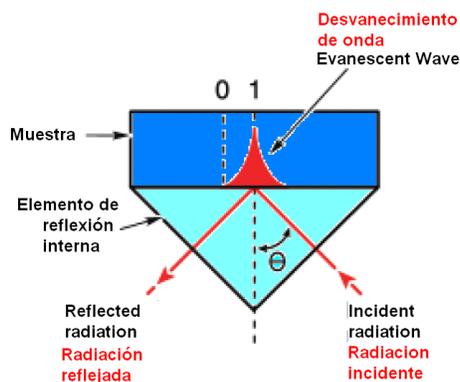


Imagen 18. Principio de funcionamiento del ATR.

La imagen 18 muestra con más detalle el principio de funcionamiento del ATR. El cristal o diamante sobre el cual se coloca la muestra está diseñado para permitir una reflexión interna total que crea una onda evanescente sobre la superficie. Esta onda se extiende a la muestra que se mantiene en contacto íntimo con el diamante, registrándose el espectro de infrarrojo del analito [27]. Este aditamento es de los más utilizados en la espectroscopia IR debido a que no es necesario el uso de porta muestras y la radiación proveniente de la fuente incide de manera directa sobre la muestra.

5.3 Quimiometría.

La quimiometría es una disciplina metrológica que enfoca su estudio en métodos matemáticos y estadísticos aplicados al análisis químico; ayuda al diseño y selección de procedimientos y experimentos óptimos proporcionando información química de calidad mediante el análisis de

datos, extrayendo la mayor cantidad posible de información [23]. Por otro lado, citando a D. L. Massart define la quimiometría como “la disciplina química que utiliza la Matemática, la Estadística y la lógica formal para diseñar o seleccionar procedimientos experimentales óptimos, proporcionar la máxima información química relevante a partir del análisis de datos químicos, y obtener conocimiento a partir de sistemas químicos” (1997) [22]. Puede considerarse como una rama especializada de la Química Analítica y la Estadística aplicada.

Su desarrollo como disciplina científica independiente se dio en el año de 1969 con los trabajos de Kowalski e Isenhour [22], quienes publicaron una serie de trabajos sobre la aplicación de máquinas de aprendizaje lineal para clasificar espectros de masa de baja resolución [34]. Con esta iniciativa surgen algunos otros grupos que basan sus trabajos en algoritmos de clasificación como SIMCA. [34].

La quimiometría como campo interdisciplinario involucra además de la estadística y química analítica otros campos como la informática y los métodos de análisis multivariable; lo que le da diversas áreas de aplicación con distintos objetivos como 1. Calibración y validación, 2. Construcción de modelos para reconocimiento de pautas y 3. Optimización de mediciones y procedimientos experimentales [35].

Dentro de este trabajo fueron utilizadas diversas técnicas quimiométricas aplicadas para diferentes fines, dichas técnicas se pueden englobar en dos grupos principales los cuales son:

1. Técnicas de filtrado de señales (Pretratamiento de datos).
2. Métodos de análisis multivariable (Técnicas que establecen agrupaciones de muestras en función de su similitud y técnicas de calibración que buscan una relación cuantitativa entre la señal analítica y una propiedad de la muestra).

En las siguientes páginas se describe con más detalle las técnicas y principios utilizados por esta disciplina.

5.3.1 Principios de quimiometría.

Para el entendimiento de las técnicas quimiométricas es necesario abordar principios básicos de estadística, por esto, empezaremos introduciendo el concepto de las diferentes variables que se estudian en este ramo, de las cuales podemos destacar dos: las “discretas” que pueden adquirir solo un número finito de valores y las “continuas” que pueden adquirir cualquier valor dentro de un intervalo [23]. De igual forma las variables también se pueden o no expresar en términos numéricos, en este sentido tenemos las variables “cuantitativas” aquellas que se mueven dentro de una escala numérica de valores y “cualitativas” o “categóricas” que no se establecen dentro de un intervalo definido de valores lo que hace que no se pueda establecer relación entre ellas [23].

Dentro de la estadística encontramos subgrupos dedicados a un análisis específico de cierto tipo de datos. La primera fase de cualquier análisis estadístico es resumir la información, de esta forma es posible dar una descripción de la muestra. La *estadística descriptiva* sintetiza la información aportada por los datos, esto lo logra estudiando los diferentes caracteres o aspectos (variables) de un conjunto concreto y resume esta información en forma de tablas, gráficos o valores [23].

En resumen entonces, la estadística descriptiva es una de las partes que compone la estadística básica. La segunda parte es la *inferencia estadística* que es la que permite tomar decisiones y establecer conclusiones posteriores sobre la población a partir de la muestra [23].

Procesamiento de datos.

El procesamiento es una parte importante cuando se trabaja con bases de datos analíticos; y más cuando se trata de datos obtenidos por medio de métodos espectroscópicos, en los cuales obtenemos gran cantidad de datos [36]. El procesamiento por utilizar depende del tipo de datos, su naturaleza y el uso que se les vaya a dar [30, 36].

Pretratamiento.

Éste se realiza antes de iniciar la construcción del o los modelos y puede ser igual o independiente del método utilizado (supervisado o no supervisado) y consiste en definir una técnica de manipulación matemática que permita remover o reducir fuentes de variación irrelevantes [30]. La tabla 5 muestra de manera general algunas de las técnicas más utilizadas.

Tabla 5. Herramientas de procesamiento de muestras.

<i>Método</i>	<i>Uso</i>
<i>Normalización</i>	Pone a todas las muestras en una misma escala dividiéndolas por una constante
<i>Ponderación</i>	La ponderación da a algunas muestras más influencia en el análisis que otras
<i>Suavizado</i>	Reducción de variaciones aleatorias (ruido)
<i>Corrección de línea base</i>	Reducir variación sistemática
<i>Centrado a la media</i>	Modelar las variaciones con respecto a la media
<i>Autoescalado</i>	Para variables predictoras de naturaleza o escala diferente
<i>Regresión Cuadrática</i>	Eliminar tendencia de espectros

Normalización.

La normalización de las muestras en la matriz se lleva a cabo dividiendo cada una de las variables entre una constante; para esto, existen dos métodos diferentes: “1-norm” y “2-norm” [30].

El primero es una normalización de área que se logra con la división de cada elemento de la matriz entre la sumatoria de cada valor “j” del vector “X”.

$$1 - norm = \sum_{j=1}^{nvars} |x_j|$$

El segundo es una normalización de longitud en la que se divide cada valor por la raíz cuadrada de la sumatoria al cuadrado de cada valor del vector.

$$2 - norm = \sqrt{\sum_{j=1}^{nvars} x_j^2}$$

De esta manera, la normalización es utilizada para remover variación sistemática muchas veces asociada al número total de muestras [30].

Ponderación.

Para el tratamiento de las muestras por este método es necesario multiplicar cada elemento de la matriz por una constante, de esta manera la ponderación es similar a la normalización, pero con un criterio diferente para definir la constante y la influencia de las muestras está definida y puede ser manipulada por un modelo matemático [30].

Suavizado.

En este método se ajusta un pequeño intervalo de longitudes de número de onda a un polinomio del grado adecuado. Los nuevos valores obtenidos después del ajuste son una mejor estimación de los valores medidos ya que se elimina la mayor parte del ruido que los afectaba. La técnica más utilizada de suavizado es la propuesta por Savitzky y Golay [30, 36].

Corrección de línea base.

En los espectros de infrarrojo no sólo existen fuentes de variabilidad de alta frecuencia (ruido), si no también algunas de baja frecuencia que no están relacionadas con la propiedad de interés, a las cuales se les denomina variaciones de línea base. Dentro de los métodos utilizados para disminuir este tipo de variaciones se encuentran el uso de 1) funciones polinómicas y 2) derivadas [36].

La imagen 19 muestra el resultado de corrección de línea base al utilizar un polinomio de grado 1.

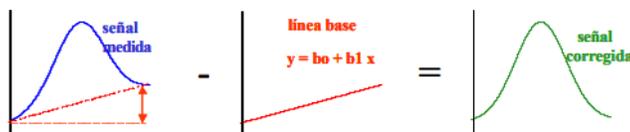


Imagen 19. Proceso de corrección de línea base utilizando un polinomio de grado 1.

Por otro lado, el uso de derivadas permite diferenciar mejor los picos solapados, además de reducir la correlación entre variables y el efecto de dispersión debido al tamaño de partículas; sin embargo, al utilizar este tipo de técnicas se debe de utilizar en conjunto con métodos de suavizado ya que se amplifica el ruido de las señales (imagen 20) [36].

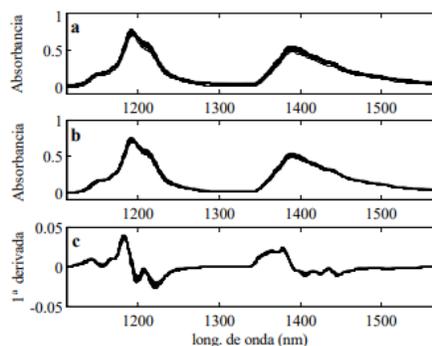


Imagen 20. Comparación entre un espectro NIR sin tratamiento (a), corrección con polinomio de grado 0 (b) y primera derivada (c).

Centrado a la media.

Técnica sencilla que consiste en restar el valor medio a cada uno de los datos espectrales; la figura 21 muestra el resultado de aplicar la técnica de centrado a un conjunto de datos (consultar sección 9.2 en los anexos) [36].

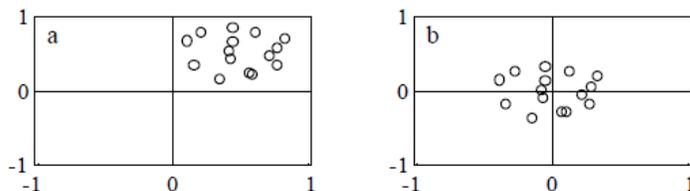


Imagen 21. a) Objetos en un plano de dos variables. b) Mismos objetos después de aplicar una técnica de centrado.

$$X'_{i,m} = X_{i,m} - X_m$$

$X'_{i,m}$: Dato centrado.

$X_{i,m}$: Dato del espectro i en la columna m

X_m : Media de la columna m

Autoescalado.

Técnica utilizada posteriormente al centrado de cada columna, se divide el resultado por la desviación estándar de la columna.

$$X'_{i,m} = \frac{X_{i,m} - X_m}{S_m}$$

$X'_{i,m}$: Dato autoescalado.

$X_{i,m}$: Dato del espectro i en la columna m

X_m : Media de la columna m

S_m : Desviación estándar de la columna m .

5.3.2 Análisis Clasificadorio y Modelado.

En un análisis clasificadorio se construyen modelos en los cuales se predice la pertenencia de un objeto (muestra) a una categoría con base en las características químicas del objeto [36].

Siendo así, para la construcción de la matriz de datos es importante considerar al menos una variable categórica, que indica la categoría a la que pertenece cada objeto y establece la variable o respuesta que se quiere predecir, y también variables de escala que describen las características de los objetos y se utilizan como variables predictoras [36].

El análisis clasificadorio es de las técnicas más utilizadas por los químicos ya que permite identificar relaciones entre objetos químicamente caracterizados; dentro de los métodos de reconocimiento de pautas o patrones podemos tener dos diferentes modalidades que dependen del conocimiento previo que se tenga del sistema [36]. Cuando contamos con un conjunto de datos con objetos que pertenecen a diferentes categorías, podemos establecer un modelo de clasificación que permita clasificar futuros objetos desconocidos dentro de una de esas categorías (métodos supervisados). Si no contamos con esta información de la categoría a la que pertenece

ninguna de las muestras, es necesario utilizar una técnica de reconocimiento de pautas que nos ayude a encontrar entre los datos grupos de composición o comportamiento similares (métodos no supervisados) [36].

5.3.3 Métodos quimiométricos no supervisados.

Un gran conjunto de datos analíticos puede ser tratado por diferentes técnicas con el fin de identificar las mejores condiciones de trabajo y las categorías por las cuales se organizan los objetos; a este tipo de métodos se les conoce como exploratorios. En ellos no se conoce o se ignora deliberadamente la existencia de categorías.

Dentro de estos métodos podemos señalar el análisis de grupos, métodos basados en redes neuronales y los basados en “*eigenvectores*” [21, 36].

Análisis de componentes principales (PCA).

Cuando se trabaja con un gran número de datos es fundamental el uso de técnicas que permitan optimizar las variables, PCA (por sus siglas en inglés: *Principal Component Analysis*) es un método matemático que permite la interpretación de las variables sin pérdida de información [2, 21]. Esta herramienta pone de manifiesto las relaciones existentes entre las diferentes muestras y en la reducción de datos, buscando las direcciones que expliquen la máxima variabilidad de las muestras (ejes ortogonales) [2, 21].

Podemos decir entonces que la información proveniente de las variables originales es proyectada a un número menor de variables subyacentes llamadas componentes principales (PC's) en donde dichas componentes están ordenadas de acuerdo con la máxima varianza [2, 20]; siendo entonces el PC1 el de máxima variabilidad, el PC2 perpendicular al primero, y de variabilidad menor pero máxima una vez restada a la variabilidad del PC1, y así sucesivamente [21].

El gráfico por el cual se proyectan estos ejes ortogonales es llamado “Gráfico de Scores” y en éste se pueden observar las relaciones entre los componentes principales, las relaciones entre diferentes variables, identificar e interpretar patrones en las muestras, diferentes agrupaciones algunas similitudes y diferencias [20].

Parte importante de un análisis PCA es el estudio de los “*loadings*”. Ya que mientras los scores nos sirven para observar las variaciones en los datos en un número menor de dimensiones; los loadings son útiles para determinar las variables originales más importantes, su estudio visual permite identificar zonas espectrales, determinando así, los compuestos químicos de mayor importancia en los datos [21].

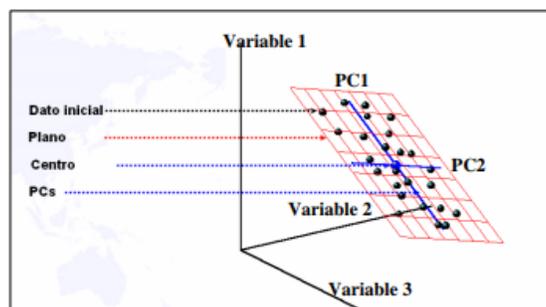


Imagen 22. Representación explicada de un diagrama de scores en 3 dimensiones.

Análisis Cluster (CA).

El análisis cluster o de conglomerados es una técnica exploratoria de clasificación complementaria a PCA. Y también al igual que el PCA divide un grupo de objetos, pero en este caso los objetos forman parte de una serie de clases en las que el grado de asociación depende de la similitud entre miembros de la misma clase [2, 21]. De esta forma el análisis de conglomerados busca objetos que se encuentren próximos en el espacio de las variables [21].

Un análisis de conglomerados trabaja con disimilitudes y métodos jerarquizados y no jerarquizados, conceptos claves que debemos de manejar para la construcción de modelos y su interpretación. Cuando nos referimos a disimilitud hablamos de una cantidad que asocia objetos (variables o clusters) parecidos con base en una matriz. Las medidas de disimilitud se realizan mediante el cálculo de distancias entre los objetos o variables [21]. Estas distancias pueden calcularse de diferentes formas [21]:

- Distancia Euclídea
- Distancia Manhattan
- Distancia de Mahalanobis

Los métodos no jerarquizados buscan la mejor agrupación posible optimizando alguna función; se basa en la partición de los grupos y no es posible establecer ninguna jerarquía entre las diferentes particiones [21]. Los métodos jerárquicos se producen por sucesión donde las particiones para los posibles agrupamientos encajan [21]. Estos métodos jerárquicos se dividen en [23]:

- Acumulativo: Trabaja bajo el hecho fundamental de que cada objeto es un conglomerado independiente al inicio, y los objetos se van uniendo por características en común en conglomerados mayores hasta formar un conglomerado único que incluye la totalidad de objetos.
- Disgregativo: Al contrario de su contraparte, este método comienza reuniendo los objetos en un único conglomerado, dividiéndolo a partir de ahí, teniendo al final tantos conglomerados como objetos.

La forma típica de representar los datos en un análisis cluster jerárquico o jerarquizado es mediante un dendrograma que consiste en la elaboración de una clasificación donde los grupos grandes se dividen en otros más pequeños [21]. Muestra los conglomerados (Clusters) que se combinan y el valor de los coeficientes de asociación en cada paso (imagen x3) [23].

Para trazar el dendrograma los conglomerados se representan por líneas horizontales o verticales proyectadas de izquierda a derecha o de arriba abajo; estas se conectan y se unen secuencialmente a su nivel correspondiente de disimilaridad conforme se cierran los clusters. Cuanto antes se cierran los clusters más parecidos entre sí son los elementos, y cuanto más tardan en cerrarse es que son menos afines [23].

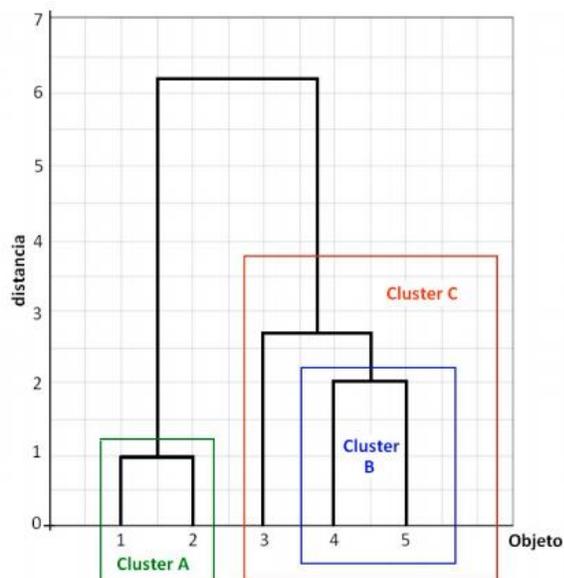


Imagen 23. Ejemplo de Dendrograma, donde se observan 3 clusters.

5.3.4 Métodos quimiométricos supervisados.

Llamamos métodos supervisados a aquellos en los que el principal objetivo es asignar a los objetos en distintas categorías con base en sus características. La construcción de modelos se realiza partiendo de una matriz de datos que contiene variables tanto categóricas (indican la categoría a la que pertenece el objeto) y de escala que son las que describen al objeto y se utilizan como variables predictoras [36].

Una de las principales características de estas técnicas es que se cuenta con dos conjuntos [36]:

- Conjunto de entrenamiento o calibración:

Muestra de objetos en la cual la categoría y valores de las variables predictoras es conocido. En este conjunto todos los objetos pertenecen a alguna categoría y de igual manera ninguno puede pertenecer a más de una [30]. Partiendo del conjunto de entrenamiento se construye el modelo de clasificación que se utiliza para predecir la categoría de nuevos objetos a partir de las medidas de la o las variables predictoras. La construcción de un buen modelo debe cumplir con dos condiciones; 1.- que el conjunto de entrenamiento cuente por lo menos con un mínimo de 4 a 5 objetos de cada categoría y 2.- las variables predictoras tienen que contener información discriminante (deben ser capaces de distinguir entre categorías) [30, 36].

- Conjunto de validación:

Objetos o muestras problema con las cuales se establece las capacidades de reconocimiento (varianza explicada de calibración), predicción (varianza explicada de validación) y robustez del modelo de clasificación. Para evaluar la capacidad de reconocimiento del modelo se toma el porcentaje de objetos del conjunto de entrenamiento que son clasificados correctamente; para la capacidad de predicción se tiene que definir previamente un conjunto de pruebas con objetos de los de los cuales se conoce su pertenencia a las categorías, es entonces que se puede definir la

capacidad de predicción como el porcentaje de objetos del conjunto de evaluación clasificados correctamente por el modelo [36]. Los objetos que constituyen el conjunto de evaluación no pueden utilizarse en la construcción del modelo de clasificación.

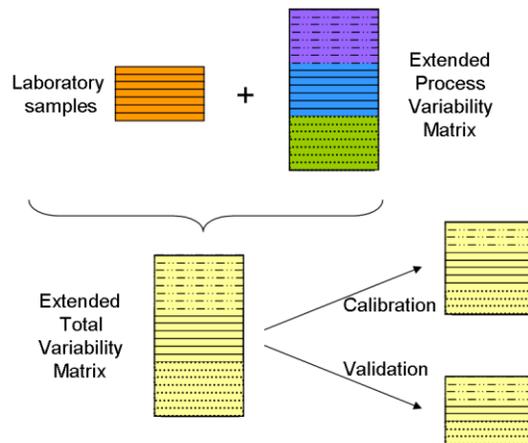


Imagen 24. Esquema para obtención de Matriz de Validación y Calibración.

Curvas de Potencia (PC).

Técnica de análisis discriminante cuadrático (ADC) utilizada en datos con categorías que no son linealmente separables o para conjuntos grandes de datos. Su representación gráfica es por medio de curvas de probabilidad elípticas o hipérbolas pertenecientes a distintas categorías [22].

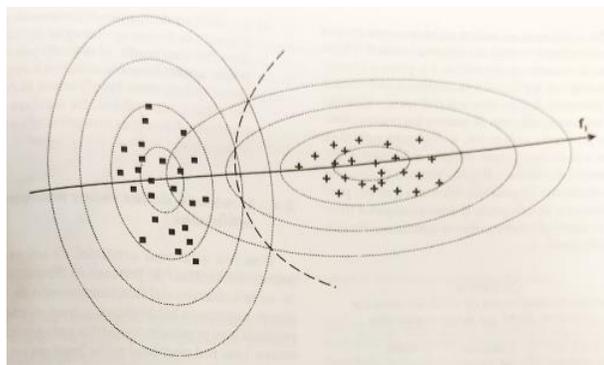


Imagen 25. Curvas elípticas entorno a los centroides de dos categorías.

Tal como se ve en la imagen, cada categoría se modela entorno a un centroide mediante una distribución de probabilidad la cual decrece en todas las direcciones del espacio m -dimensional de las variables. Esta probabilidad aumenta en los puntos situados cerca del centroide de cada categoría y baja para los puntos más alejados [22].

La clasificación de objetos mediante esta técnica se hace en función de sus probabilidades de pertenencia a una de las categorías; de tal forma que el objeto se asigna a la categoría a la que tiene mayor probabilidad de pertenencia, existiendo así muestras que tengan cierto porcentaje de pertenencia a más de una categoría [22. 25].

Se denomina frontera o región crítica a la zona que separa las categorías, en este hiper-plano se igualan las probabilidades de pertenencia de un objeto a una u otra categoría [22].

Funciones de Potencia (PF).

Este método puede utilizarse con dos fines, ya sea para clasificar o como modelado de clases. Suele ser muy eficiente para datos que no responden a una distribución normal al realizar un tratamiento de datos. A diferencia de otros métodos en los cuales se estima la función de densidad de probabilidad partiendo de la media y desviación estándar, para los modelos construidos bajo esta técnica se utiliza la densidad local de objetos (suma de contribución individual) en el conjunto de calibración [23].

A cada contribución individual le corresponde una función Gaussiana que contiene un parámetro de suavizado “ u ” que determina el ancho de cada contribución individual. La distribución de probabilidad de cada punto o muestra x en la matriz de datos es la suma de contribuciones individuales; mientras que para la elección del parámetro de suavizado se debe de realizar un procedimiento previo de optimización [22, 23].

Cuando hablamos de datos multivariantes, el potencial individual en un punto x debe tener en cuenta las V variables existentes. El parámetro de suavizado está dado por un factor de suavizado k que se calcula con la desviación estándar de cada variable, de forma que solo el parámetro k debe ser optimizado. Si se estudian diversas categorías o clases, el factor de suavizado k se calcula para cada una de ellas [23].

En el programa Matlab este procedimiento de optimización para obtener el factor de suavizado consiste en la obtención de diversos gráficos por cada una de las clases de estudio en el que se determina el error, sensibilidad y especificidad.

Vecino más cercano (K-NN).

K-NN es una de las técnicas que se encuentra dentro de los métodos quimiométricos de clasificación no paramétricos de enlace simple, es decir, no se basa en un modelo; calcula la distancia de aproximación más simple del objeto nuevo (test) a los objetos del conjunto de entrenamiento [22, 23], construyendo de esta forma una matriz de disimilaridades [23]. De los algoritmos más utilizados de esta técnica está k -NN o “Vecinos próximos”, que clasifica al objeto nuevo en la categoría a la que pertenecen sus k vecinos más próximos, donde k es un valor arbitrario [22].

El método calcula las distancias entre el objeto nuevo y los otros objetos del grupo o conjunto de entrenamiento.

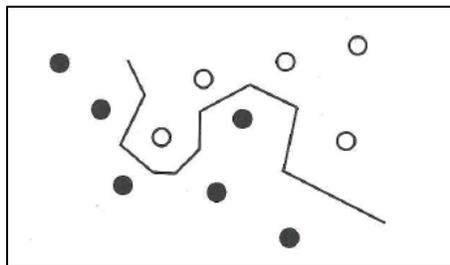


Imagen 26. Límite de separación para la clasificación de objetos de dos clases con $k=1$.

5.3.4.1 Analogía de clases por modelación suave e independiente (SIMCA).

Además de los métodos discriminantes, es posible determinar la clase de un objeto utilizando la descripción de cada clase individual en un modelo matemático independiente de las otras clases. SIMCA por sus siglas en inglés (Soft Independent Modeling of Class Analogy) es un método de clasificación supervisado en el que cada clase se modela utilizando un PCA independiente [25].

Se puede describir dicho modelo como una “Class box” (espacio de clase) en la cual los objetos desconocidos pueden ser clasificados o ajustados acorde a un modelo óptimo y específico de cada clase en particular en el cual se encuentran las muestras que son descritas por el modelo como pertenecientes a la clase de estudio y las muestras que no se consideran pertenecientes a la clase [22, 25].

El modelo determina si una nueva muestra se ajusta a la clase mediante un cálculo matemático en el cual se calcula la varianza residual total de la clase y la varianza residual de la muestra según las fórmulas 1A y 1B [25].

1A. Varianza residual total de la clase.

$$s^2_0 = \sum_{i=1}^n \sum_{j=1}^p \frac{e^2_{ij}}{(n - A_q - 1)(p - A_q)}$$

1B. Varianza residual de la muestra.

$$s^2_i = \sum_{j=1}^p \frac{e^2_{ij}}{p - A_q}$$

Si ambos valores de varianza son del mismo orden de magnitud la muestra es asignada como un objeto típico de la clase. Pero cuando $s^2_i > s^2_0$ es eliminado. Dicho de otra forma, cuando los residuales de la muestra son significativamente mayores que los de la clase, la muestra es rechazada [25].

Para cada clase q se construye un modelo separado, según la fórmula siguiente:

$$X_{ij}^q = \bar{X}_j^q + \sum_{a=1}^{A_q} t_{ia}^q l_{ja}^q + e_{ij}^q$$

Donde:

\bar{X}_{qj} = Media de la variable j en la clase q .

A_q = Número de componentes principales significativos en la clase q .

t_{ia}^q = Scores del objeto i en el componente de la clase q .

l_{ja}^q = Loadings de la variable j en el componente principal a en la clase q .

e_{ij}^q = Error residual del objeto i y variable j .

En la figura x observamos las diferentes formas típicas de graficar un modelo por SIMCA, estos pueden ser representados de forma lineal por un componente “A_q”, en un plano por dos componentes principales o en un plano de tres dimensiones por tres componentes principales. El modelo también permite identificar las muestras atípicas u “Outliers” (muestras que no pueden ser asignadas a ninguna clase del modelo) [25].

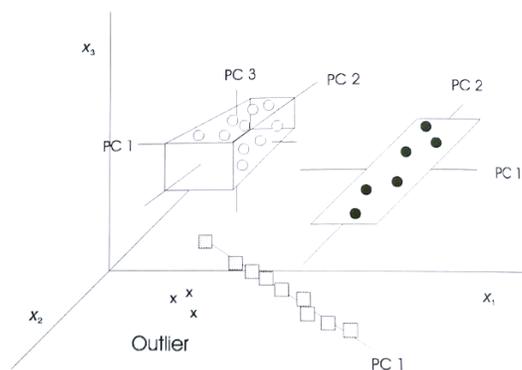


Imagen 27. Modelo de suavizado independiente.

6. Metodología.

El presente trabajo se desarrolló por la necesidad de resolver los problemas a los que se enfrenta la industria tequilera hoy en día como son la adulteración de bebidas y la caracterización de los tequilas por la procedencia de sus azúcares para la fermentación (tequila 100% y mixto).

Las metodologías desarrolladas, pueden ser divididas de dos formas; las enfocadas a la caracterización de las bandas y su comparación con los principales componentes del tequila y las enfocadas al desarrollo de modelos de clasificación.

Los modelos óptimos obtenidos podrán llevarse a cabo por el personal del laboratorio del CRT y servirán como técnica complementaria a los análisis que se realizan actualmente por esta institución.

6.1 Actividades.

6.1.1 Caracterización de bandas.

1. Obtención de espectros de transmitancia FT-MIR de muestras
2. Pretratamiento de los datos; transformación en absorbancia, corrección de ATR, corrección automática de línea base y ajuste a cero.
3. Elección de tequila bien comportado de cada una de las categorías, tequila 100% agave y tequila mixto.
4. Búsqueda de espectros de los principales componentes del tequila en librería web.
5. Comparación de espectros obtenidos con los encontrados en la librería.
6. Resumir y concluir.

6.1.2 Modelos de clasificación.

1. Obtención de espectros de transmitancia FT-MIR de las muestras de tequilas.
2. Pretratamiento de los datos; transformación en absorbancia, corrección de ATR, corrección manual de línea base y ajuste a cero.
3. Elección del intervalo de trabajo y pretratamiento de los datos.
4. A partir de los datos espectrofotométricos se realiza la construcción de las matrices y se lleva a cabo el análisis multivariante mediante la utilización de dos programas computacionales (GenEx[®] (MultiDAnalysis AB, Göteborg, Suecia) y Matlab[®] (MathWorks, US.) para la obtención de modelos óptimos para la diferenciación entre categorías (Tequila 100% agave y mixto).
5. Realizar análisis de los resultados obtenidos.
6. Comprobar la validez de los modelos con muestras de prueba (*test*) del CRT.

7. Experimentación.

7.1 Diseño práctico.

En esta sección se presentará el trabajo experimental efectuado en el laboratorio, así como la metodología utilizada para el tratamiento de los datos obtenidos. Toda la experimentación fue realizada en las instalaciones de la Unidad de Investigación Multidisciplinaria (UIM) de la FESC Laboratorio 10 de Físicoquímica Analítica.

7.1.1 Materiales, reactivos y equipos.

Materiales.

- 5 Matraces aforados de 10 mL.
- 1 Pipeta volumétrica de 0.5mL, 3mL, 4mL y 5mL.
- 1 Perilla de succión.
- 2 Vasos de precipitados de 50 mL.
- 1 Micropipeta Finnpiptette de 50 microL.
- 1 Micropipeta Biohit proline 100-1000 microL.
- Puntas para Micropipeta.
- Frasco ámbar de 500 mL.
- Papel óptico.

Reactivos.

- Etanol absoluto (100% de pureza).
- Etanol Anhidro (99.9% de pureza).

Equipo.

- Espectrofotómetro Perkin Elmer Frontier FT-IR con UATR de diamante/KRs5 (TlBr-TlI) de una reflexión.

7.1.2 Muestras de tequila.

Para la experimentación fueron utilizadas 63 muestras certificadas por el CRT, de las cuales fueron medidos 35 tequilas 100 % agave y 28 tequilas mixtos. Aunado a esto, también fueron medidas 18 muestras adicionales con diferentes proporciones de alcohol; 3 muestras con 35 % de alcohol, 1 muestra con 38 %, 11 muestras con 40 %, 1 muestra con 50 % y 1 con 55 %.

Código de las muestras.

Fue asignado un código de letras y números a las muestras medidas con el fin de diferenciar entre categorías y clases. Este código fue utilizado para diferenciar entre clases de tequila se muestra en la figura 1.

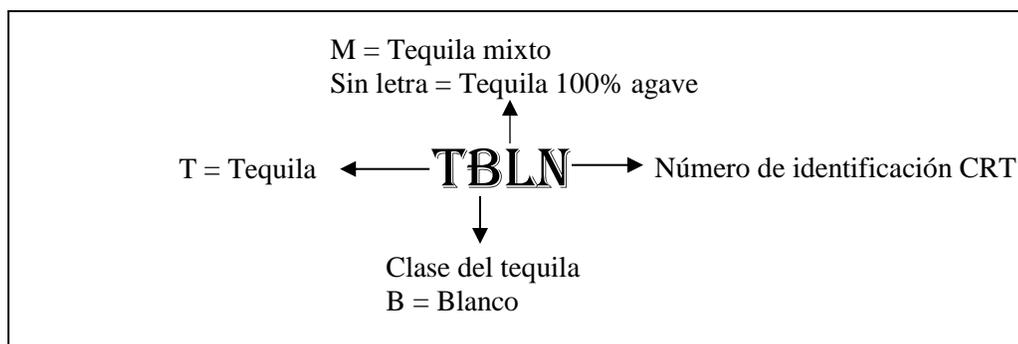


Figura 1. Nomenclatura en muestras de tequila

Para la diferenciación entre categorías se añadió una letra “M” a los tequilas mixtos, ejemplo: “TB01” (tequila blanco 100 % agave) y TBM09 (tequila blanco mixto). En la tabla 6 se muestra los símbolos utilizados para identificar los tequilas por categoría en los diferentes *softwares* utilizados.

Tabla 6. Símbolos y colores para los tequilas

Bebida	Simbología <i>GenEx</i> ©	Simbología <i>Matlab</i> ©	
		Calibración	Validación
Tequila 100% Agave			
Tequila Mixto			

7.1.3 Condiciones experimentales.

Las mediciones de los tequilas en el espectro FT-IR se hicieron corriendo un blanco entre cada muestra en un intervalo de 400 a 4000 cm^{-1} con una resolución nominal de 1 cm^{-1} .

7.1.4 Softwares.

La utilización de herramientas computacionales fue de vital importancia para el trabajo ya que todos los tratamientos y modelos obtenidos se desarrollaron con los siguientes programas.

Spectrum© (Perkin Elmer, Waltham, MA, U.S.A.)

GenEx© (MultiD Analysis AB, Göteborg, Sweden).

Matlab© (The MathWorks, Massachusetts, U.S.A.) con la herramienta Classification Toolbox (v.3.1)

Los espectros de los principales componentes del tequila se recopilaron de la Web de Química del “*National Institute of Standards and Technology*” (NIST).

7.2 Obtención de espectros FT-MIR de muestras de tequilas para su comparación y la construcción de modelos.

Elaboración y obtención de espectros de tequilas para su comparación con espectros de sus principales componentes y la construcción de modelos.

- a. Encender equipo de cómputo, entrar a programa *Spectrum*© y elegir método experimental “x” previamente cargado.
- b. Con ayuda de una piseta limpiar el diamante del ATR con etanol absoluto, secar con papel óptico dejar evaporar el resto del alcohol.
- c. Ingresar código de tequila a medir con base en el apartado 6.1.2 y correr blanco contra aire.
- d. Colocar punta de plástico en la Micropipeta Finnpiette de 50 μL y enjuagar de dos a tres veces con la muestra de tequila.
- e. Colocar sobre el diamante del ATR 50 μL de muestra y correr el barrido de 400 a 4000 cm^{-1} .
- f. Medir todas las muestras en orden aleatorio repitiendo los pasos 2 a 5.
- g. Realizar pretratamiento de datos (apartado 6.2.2.1).
- h. Construcción de matriz de trabajo.

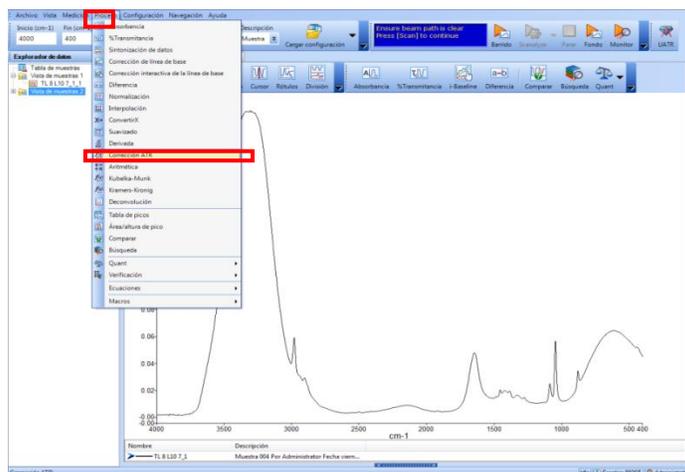
7.2.1 Pretratamiento de datos.

En este apartado se describe el proceso al cual fueron sometidos los datos experimentales antes de elaborar las matrices; el pretratamiento a los datos experimentales es fundamental al trabajar con técnicas quimiométricas ya que ayudan a la normalización de las señales y variables.

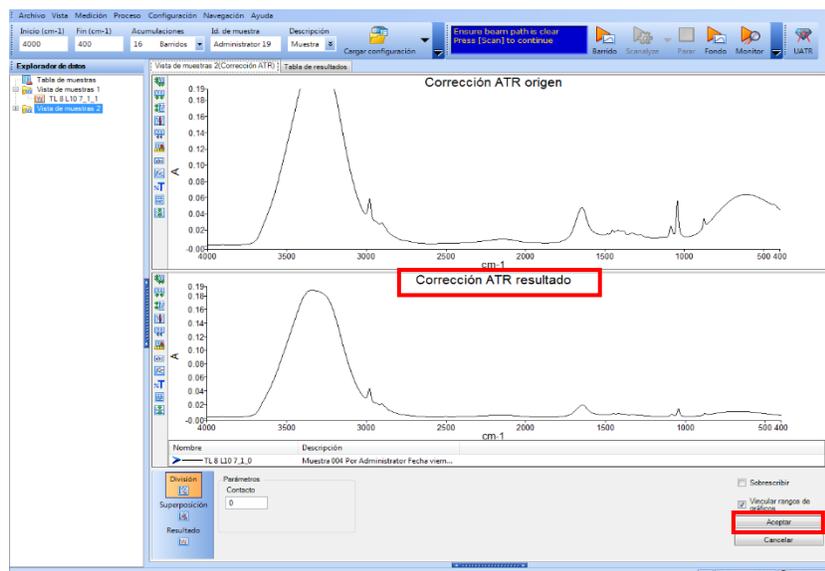
Una vez obtenidos los espectros FT-MIR de las 63 muestras en porcentaje de transmitancia; éstos fueron convertidos a absorbancia. Todo el pretratamiento de datos se realizó con el software *Spectrum*©.

Corrección de ATR.

Los espectros fueron obtenidos por reflexión así que se realizó una corrección de ATR.



Paso	Acción
1	Seleccionar todos los espectros.
2	Dar clic en la pestaña "Proceso".
3	Seleccionar "Corrección de ATR".



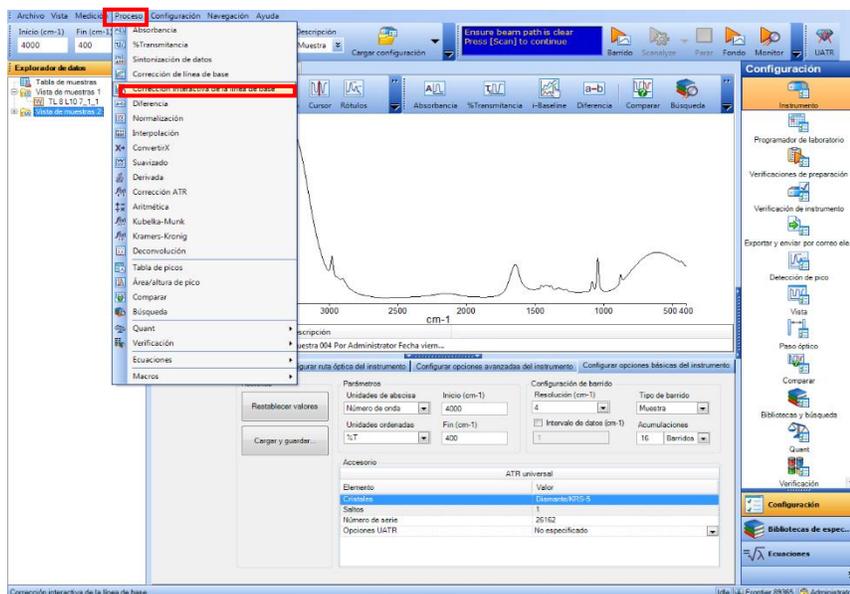
Paso	Acción
4	Se visualiza la corrección de los espectros.
5	Clic en "aceptar".

Corrección interactiva de línea base.

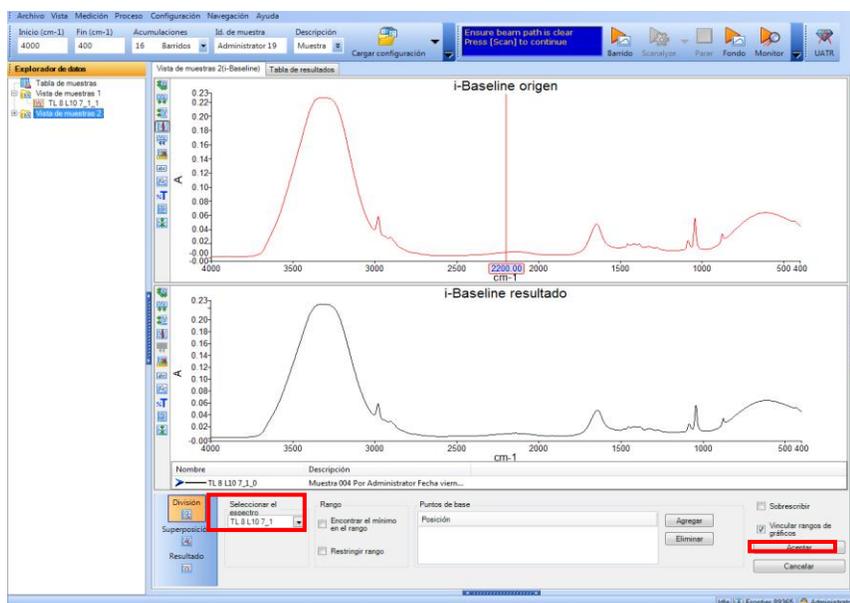
Se realizó la corrección de línea base por categoría considerando un tequila representativo de cada caso. Los puntos tomados para hacer la corrección se presentan a continuación en valores de número de onda (cm^{-1}):

Tequilas 100% agave: 4000, 3912.87, 2597.14, 1829.97, 1501.8, 1186.62, 983.2, 400.

Tequilas mixtos: 4000, 3895.82, 2581.79, 1867, 1503.03, 1174.74, 986.02, 400.



Paso	Acción
1	Seleccionar espectros corregidos por ATR.
2	Dar clic en la pestaña "Proceso"
3	Seleccionar "Corrección interactiva de línea base."



Paso	Acción
4	Elegir un espectro característico de cada categoría.
5	Corregir sobre los números de onda propuestos al inicio de la sección para cada categoría.
6	Clic en "aceptar".

Ajuste a cero.

Para el ajuste a cero se tomó el valor obtenido de cada espectro a 400 cm^{-1} y se restó a los otros valores obtenidos mediante Excel.

7.2.2 Construcción de matrices.

La construcción de la matriz de trabajo se realizó a partir de los espectros MIR de las 63 muestras de tequila blanco (35 tequilas 100 % agave y 28 tequilas mixtas) previamente tratados. A partir de ésta, se llevó a cabo la construcción de los modelos exploratorios de PCA y CA por GenEx.

Matrices de calibración y validación.

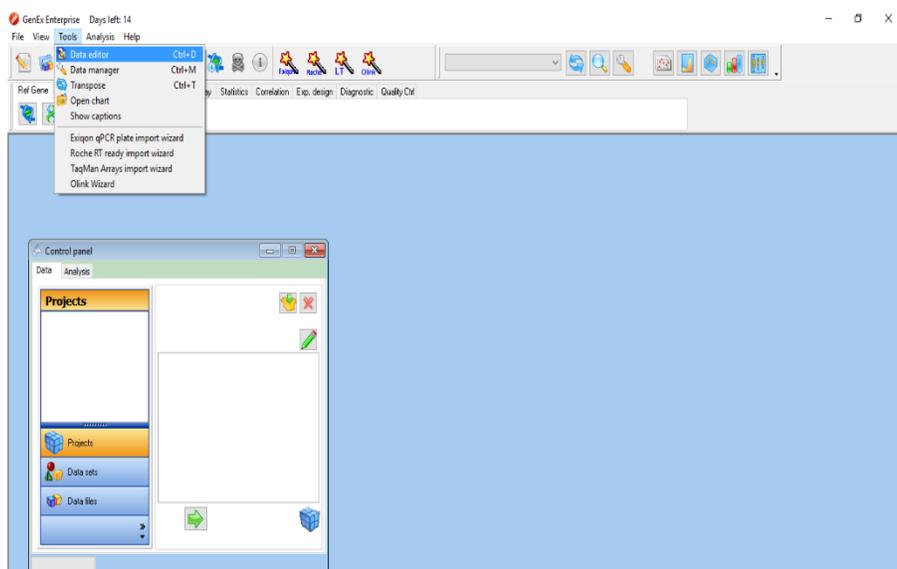
Por otra parte, tomando en cuenta los modelos de exploración, se elaboraron dos matrices, una de calibración y validación. La de calibración se realizó con los datos de MIR 52 muestras, 31 tequilas 100% agave y 21 tequilas mixtos. Para la de validación fueron seleccionadas 8 muestras test, de las cuales 4 son de tequilas 100% agave (TB91, TB92, TB101 y TB127) y 4 de tequilas mixtos (TBM10, TBM41, TBM97 y TBM131).

Fueron encontradas 3 muestras anómalas las cuales no se consideraron para los modelos obtenidos por técnicas supervisadas. Las muestras tomadas para la matriz de validación no fueron consideradas en la matriz de calibración.

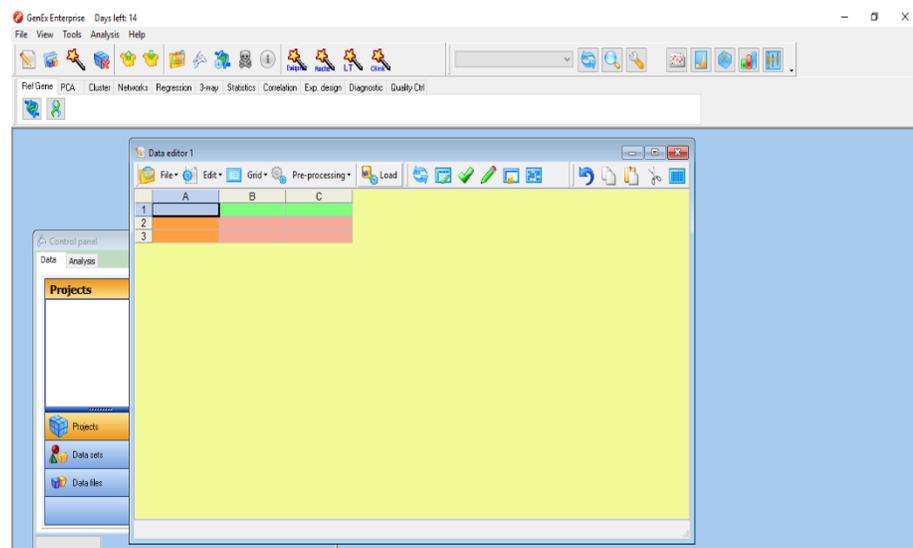
Se compararon un tequila representativo de cada categoría (100% agave y mixto) a fin de vislumbrar las principales diferencias entre ellos y con ello los posibles grupos funcionales.

Pasos para la obtención de modelos por GenEx ©.

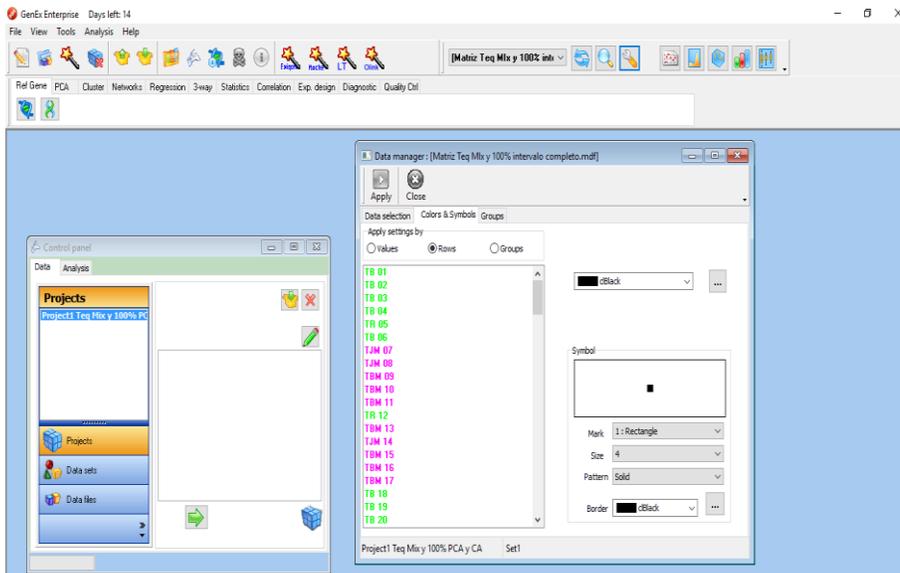
Iniciar programa, en la pantalla inicial:



Paso	Acción
1	Dar clic en la pestaña "Tools".
2	Seleccionar "Data editor".



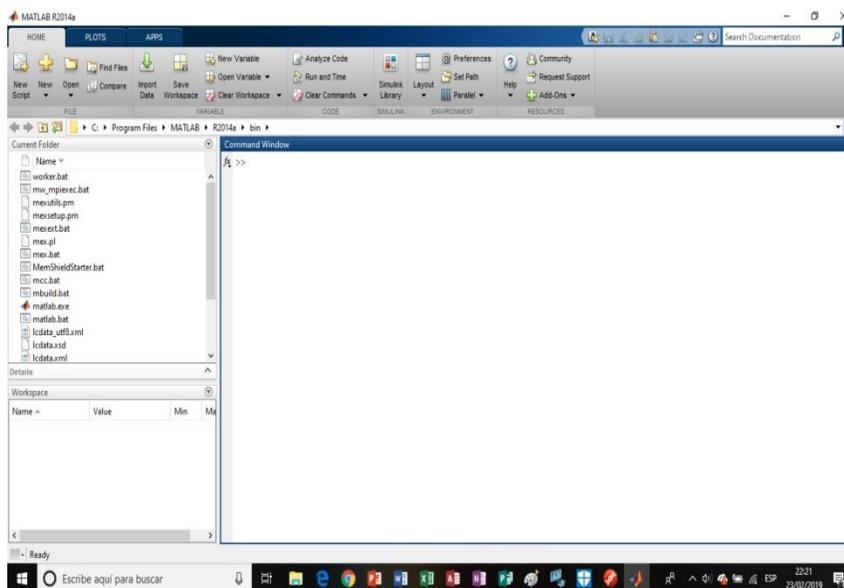
Paso	Acción
3	Pegar matriz de Excel con datos tratados.
4	Clic en "Load" y guardar matriz



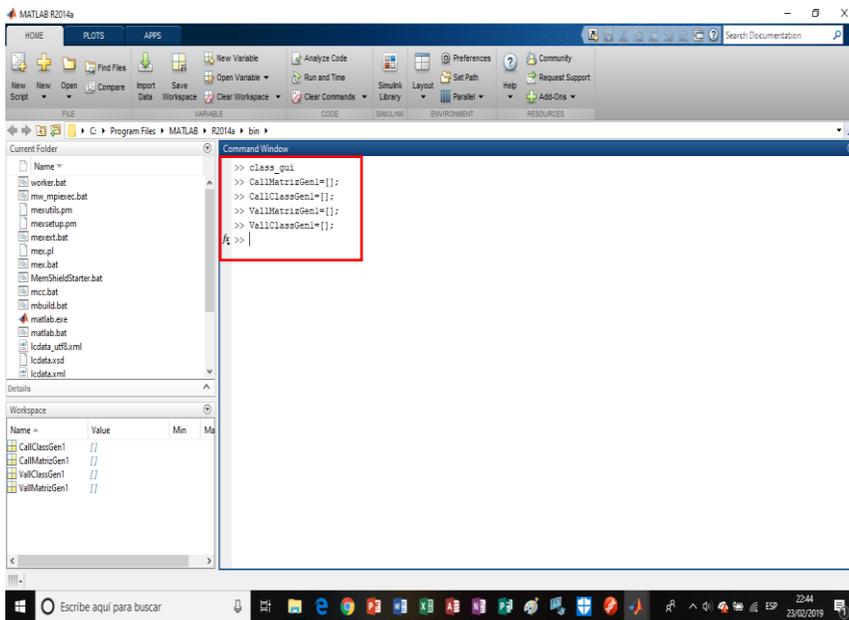
Paso	Acción
5	Clic en “Data Manager”
6	Ir a “Colors & symbols”
7	Asignar color y forma a cada categoría con base en la sección 5.1.2.
8	Dar clic en “Apply” y correr modelo.

Pasos para la obtención de modelos por Matlab ©.

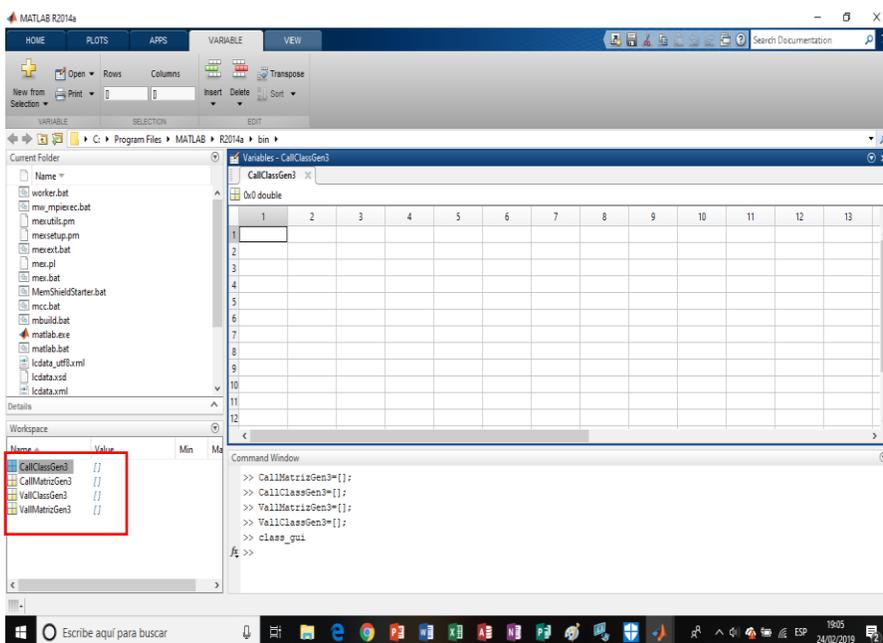
Iniciar programa, en la pantalla de inicio:



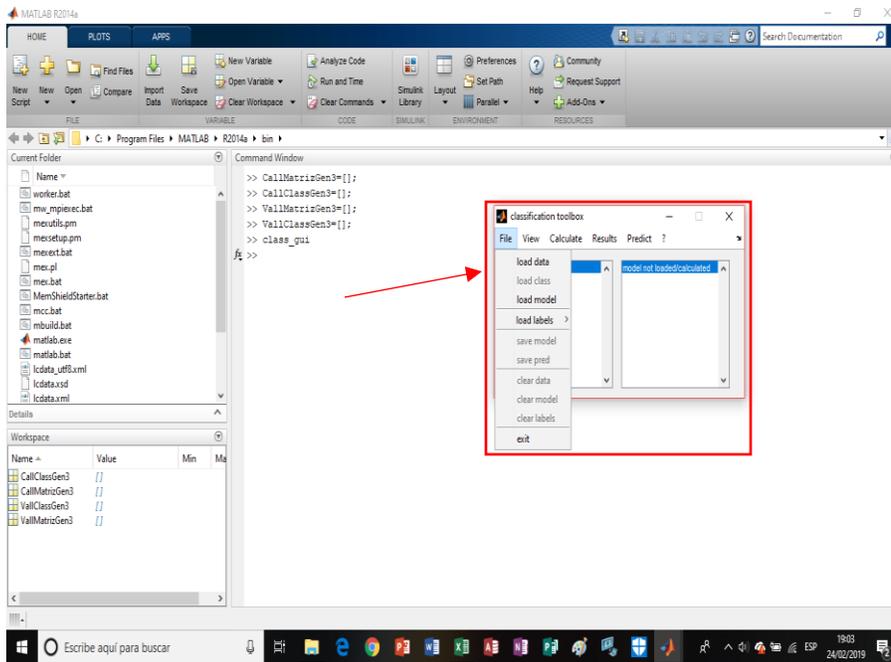
Paso	Acción
1	Posicionarse sobre la ventana de comandos
2	Escribir: “class_gui” para activar ventana de trabajo.



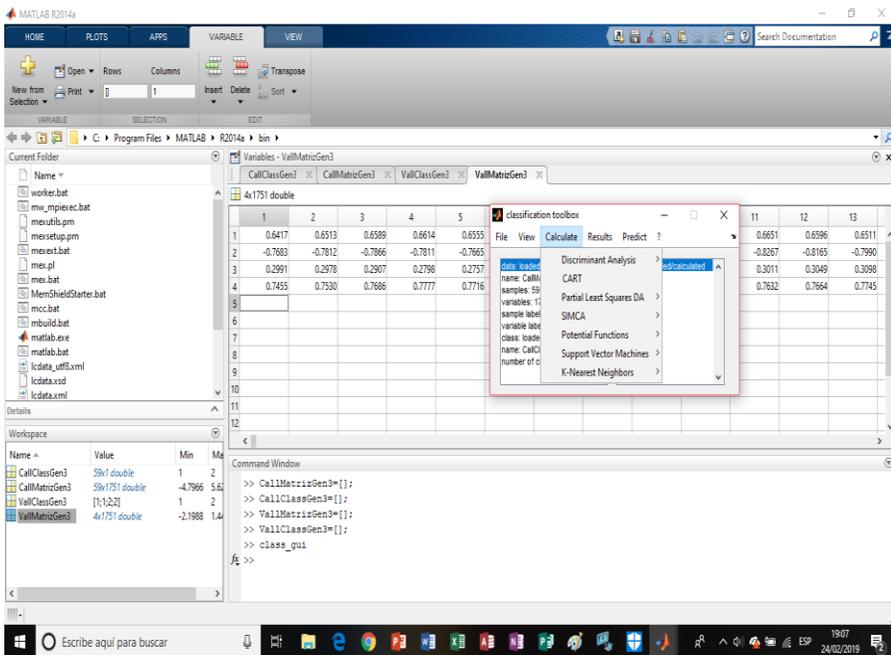
Paso	Acción
3	Activar hojas de cálculo para ingresar datos de la matriz de calibración y validación.
4	Escribir: “CallMatriz= [];” para ingresar matriz de calibración y “CallClass= [];” para la clase de calibración.
5	Escribir: “VallMatriz= [];” para ingresar matriz de validación y “VallClass= [];” para clase de validación.



Paso	Acción
6	Dar doble clic sobre los comandos ingresados en los pasos 4 y 5.
7	Al abrir la hoja de cálculo, ingresar los datos de la matriz de validación y calibración respectivamente.
8	Para ingresar la clase se asigna el número 1 a los tequilas 100% agave y el número 2 a los tequilas mixtos.



Paso	Acción
9	En la ventana de trabajo ingresar datos de matriz y clase de calibración
10	Dar clic en "Calcular" y seleccionar modelo.



Paso	Acción
11	Una vez seleccionado el modelo, calcular parámetros de optimización.
12	Correr modelo seleccionado con los parámetros óptimos.
13	Cargar datos de validación y dar clic en "Predecir".

8. Resultados y Discusión.

En este capítulo se describe un análisis comparativo realizado entre los espectros MIR de un tequila 100 % agave y mixto con el fin de identificar las zonas de mayor diferencia y con ello determinar el o los intervalos óptimos para llevar a cabo la construcción de los modelos de clasificación.

A continuación, la Figura 2 muestra los espectros de tequilas 100 % agave y mixtos de los diferentes muestras de tequila en el intervalo de trabajo 400 a 4000 cm^{-1} contra un blanco aire, por lo que su aspecto es debido a los componentes mayoritarios del tequila, agua- etanol.

A primera vista observamos en estos espectros una diferencia en la banda de 3600 a 3000 cm^{-1} correspondiente a interacciones del enlace O-H del grupo hidroxilo del agua y etanol de la matriz. El ancho de estas bandas se debe a la formación de puentes de hidrógeno. Además, en experimentos previos, se ha comprobado la intensidad de esta banda disminuye de manera proporcional al aumentar el porcentaje de etanol (figura 2).

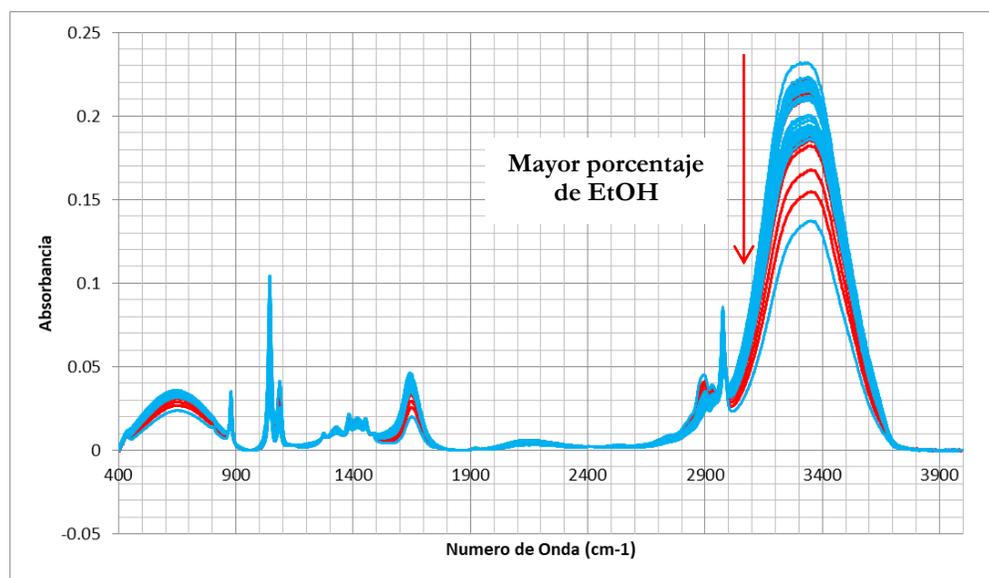


Figura 2. Espectros de las diferentes muestras de tequila en estudio.
Tequilas Mixtos en azul y Tequilas 100 % agave en rojo.

8.1 Análisis comparativo e interpretación de los espectros MIR de un tequila 100 % agave y un mixto.

En esta sección se muestra los resultados obtenidos al realizar la comparación de dos tequilas representativas de cada categoría (TB53 y TBM46) con los espectros de los principales compuestos presentes en el tequila (Anexos sección 10.4). De esta manera, se identifican los grupos funcionales correspondientes a las principales señales del espectro

El tequila, como se ha mencionado en capítulos previos, contiene múltiples compuestos químicos los cuales dependen tanto de la fuente de la materia prima utilizada, como del proceso de elaboración.

Existen diversos estudios y trabajos de investigación en los que se han identificado y cuantificado algunos de estos compuestos (37).

La figura 3 presenta los espectros de dos tequilas representativas de cada categoría. Las flechas señalan los principales grupos funcionales asociadas a las señales observadas. En la mayoría de las bandas es posible asignar más de un grupo funcional, dada la complejidad química del tequila.

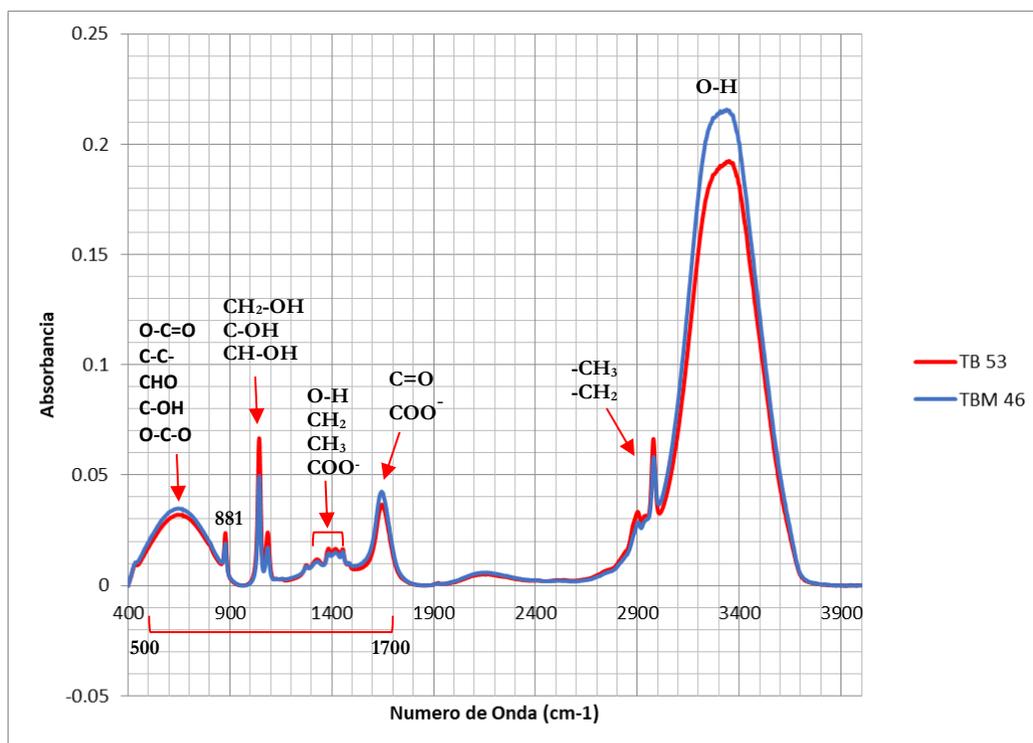


Figura 3. Espectros FT-MIR representativos de cada clase de tequila y sus principales grupos funcionales por banda. Tequila 100% agave en rojo y Tequila mixto en azul.

A primera vista, en los espectros destacan algunas señales por su intensidad. En particular, en el tequila mixto (TBM46) las señales de 3000 a 3600 cm^{-1} , 1650 cm^{-1} , 2800 a 3000 cm^{-1} y 400 a 850 cm^{-1} presentan una mayor intensidad que en el tequila 100 % agave (TB53), mientras que la intensidad de las bandas de 1277 a 1455 cm^{-1} , 1050 cm^{-1} y 900 cm^{-1} es mayor en el tequila 100 % agave que en el tequila mixto.

Como se mencionó en la sección anterior, la banda 3000 a 3600 cm^{-1} se debe a estiramientos del enlace O-H y se ensancha por la formación de puentes de hidrógeno; por otro lado, su intensidad está relacionada con el porcentaje de alcohol en el tequila por lo que se deduce que la diferencia en esta zona es debido a que el TBM46 presenta un menor contenido de etanol que TB56. Esto también se ve reflejado en la banda ancha de 400 a 850 cm^{-1} , típica de los alcoholes de cadena corta (estiramientos del enlace C-O).

Las diferencias en la intensidad de estas bandas entre los tequilas 100 % agave y mixtos puede verse influenciada por la cantidad de agave utilizado, ya que una mayor cantidad de agave propicia el aumento de pectinas disponibles, las cuales están relacionadas directamente con la cantidad de metanol (37).

Los componentes del tequila que se utilizaron para la comparación con los espectros fueron los que de acuerdo con la literatura se presentan en mayor concentración. Dichos compuestos fueron: furfurales (Furfuraldehído, 2-hidroxi metil furfural, 2 y 5 metilfurfural, acetato de 2-Furanmetanol y 2,5-Dimetilfurano), Benzaldehído, Etil éster y Acetaldehído. Estos compuestos son los principales responsables del perfil espectroscópico de los tequilas generando bandas principalmente en la zona de 500 a 1700 cm^{-1} (37). Otro componente tomado en cuenta fue el Oxalato, debido a que diversos estudios lo colocan como un compuesto abundante y presente en los tejidos del agave *tequilana Weber* (38); además, en un análisis cromatográfico este compuesto fue usado para diferenciar tequilas por categoría (37), por lo que también fue considerado para la presente clasificación de los tequilas.

La tabla 7 muestra las principales señales en el intervalo 500 a 1700 cm^{-1} explicadas por grupo y por vibración del enlace. Dentro de las zonas más destacadas, está la de 1300 a 1450 cm^{-1} la cual incluye a grupos funcionales de compuestos tales como el etil éster cuyas señales corresponden a estiramientos simétricos del grupo COO. En este grupo de bandas también observamos contribución de los grupos alifáticos presentes en el acetaldehído, compuestos furfurales (CH_3) y de vibraciones de torsión/estiramiento del carboxilo en compuestos como el ácido carboxílico. De las principales contribuciones del benzaldehído en este tipo de espectros, observamos la banda de 881 cm^{-1} , ocasionada por deformaciones fuera del plano del C-H en el anillo de benceno, así como las bandas por interacciones del tipo CH=C que se presentan en alquenos y anillos aromáticos sustituidos (carbono en posición alfa al aldehído) en la zona de 1485 a 1615 cm^{-1} .

Otro intervalo en el que observamos contribuciones de los compuestos furfurales y el acetaldehído es el ubicado entre 600 y 880 cm^{-1} que corresponde a desdoblamientos del enlace C-C en el grupo C-CHO del aldehído. De igual forma, el heteroátomo de oxígeno en la molécula de furano y sus interacciones con los carbonos del ciclo también contribuyen a la formación de las bandas dentro de esta zona.

Una de las señales más características en el tequila que es la observada en 1050 cm^{-1} debida principalmente al etil éster por estiramientos simétricos y asimétricos de los enlaces C-O-C y O-C=O, compuestos furfurales y en menor proporción estiramientos del enlace C-OH en alcoholes.

Por último, la señal de 1600 cm^{-1} típicamente intensa y la cual es asignada al grupo carbonilo, en este caso se debe agua principal componente de la matriz del tequila (39). Efectivamente, al analizar el espectro del agua (figura 4) se observa la misma señal, con una intensidad similar, por lo que podemos deducir que esta señal en realidad corresponde al agua y no al carbonilo, o bien existe un traslape en la donde predomina la intensidad del agua.

Tabla 7. Señales características del tequila y sus grupos relacionados intervalo 500 a 1700 cm^{-1} .

Grupo	Longitud de Onda	Enlace				Tipo de compuestos
		C-H	C-C	C-O	O-H	
CH ₂	1475-1450	Flexión de tijera				Compuestos alifáticos
CH ₃	1465-1440	Deformación asimétrica				Compuestos alifáticos
	1380-1370	Deformación simétrica				Isopropilo
	1380-1360	Deformación simétrica				
CH ₂ -OH	1060-1025			Estiramiento		Alcoholes primarios
C-H	850-790	Deformación fuera del plano				Alquenos
COO ⁻	1400-1310			Estiramiento simétrico		Sales derivadas de ácidos carboxílicos
	1610-1560			Estiramiento asimétrico		
Ar	1515-1485		Estiramiento del anillo enlace CH=C			Benceno
	1615-1590					
	830-740	Deformación fuera del plano				Anillo de benceno sustituido
	720-600				Deformación fuera del plano	Fenoles
OH	1440-1400				Desdoblamiento en el plano	Ácidos carboxílicos
C-O-C	1280-1150			Estiramiento asimétrico		Esteres
	645-575			Desdoblamiento		
C-OH	1200-1015			Estiramiento		Alcoholes (general)
	1120-1080					Alcoholes secundarios y terciarios
	680-620			Desdoblamiento		Alcoholes (general)
C-CHO	695-635		Desdoblamiento			Aldehídos
	565-520			Desdoblamiento		
O-C=O	700-590			Desdoblamiento		Ácidos carboxílicos

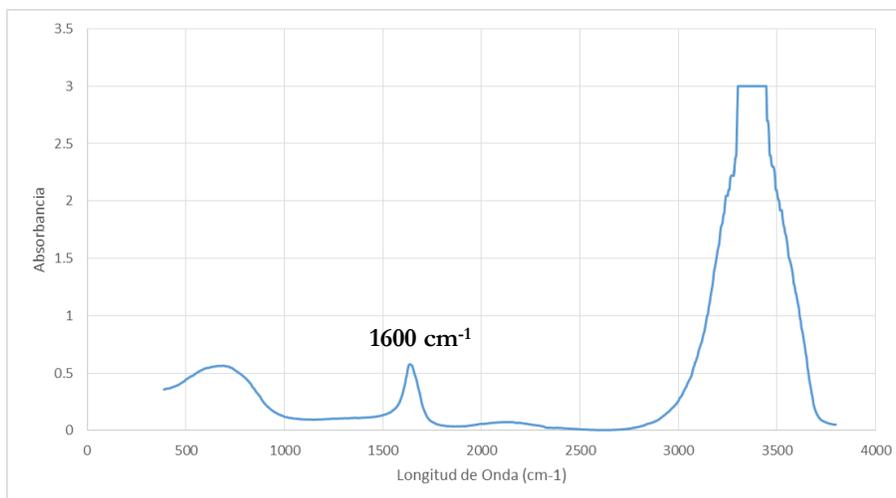


Figura 4. Espectro FT-MIR del agua.

Como hemos visto hasta ahora, el perfil del espectro del tequila es el resultado de numerosos compuestos, por lo que en este apartado sólo hemos mencionado los grupos funcionales de los compuestos en mayor proporción.

Para la elaboración de los modelos de clasificación es determinante elegir el intervalo óptimo. En el presente estudio se tomó en cuenta las zonas donde los compuestos más representativos están presentes, así como aquellas donde la diferencia entre ambas categorías es máxima. En sentido, el intervalo elegido fue de 400 a 1600 cm^{-1} donde las señales C-O, C-H son más específicas y que al mismo tiempo no incluyeran las señales O-H de la matriz.

8.2 Modelos exploratorios con técnicas no supervisadas.

8.2.1 Análisis de componentes principales (PCA).

Como se menciona en el apartado de experimentación, la corrección de la línea base se realizó de manera interactiva. Para cada categoría de tequila se seleccionaron valores de número de onda, de manera de evitar valores negativos de Absorbancia. Posteriormente se hizo un ajuste a cero.

En un trabajo paralelo se realizan correcciones de línea base con diferente criterio ya que se ha descubierto que este criterio es esencial en la construcción de los modelos

Previo a la realización del modelo inicial, se retiró de la matriz el tequila TBM49 ya que con base en información del CRT está identificado como licor de agave, y para los modelos solo se trabajó con tequilas blancos.

El modelo preliminar se realizó en todo el intervalo (400 a 4000 cm^{-1}), los datos utilizados únicamente presentan corrección de línea base, ajuste a cero y sin escalado; el resultado de este primer modelo se presenta en la figura 5 donde se puede observar una distribución de los scores de los tequilas a lo largo del subespacio PC1-PC2 sin una clara separación entre las categorías. Tres PCs explican el 99.99 % de la varianza acumulada, donde PC1 contribuye con un 99.87 %, PC2 0.12 % y PC3 un 0.0021 %. La representación en tres dimensiones permite vislumbrar la formación de dos grupos que contienen la mayoría de las muestras (1A y 2A), los cuales están

compuestos por tequilas de ambas categorías. Además, se forman otros dos pequeños agrupamientos. El subgrupo 1B contiene tres tequilas: TBM03, TBM17 y TBM23 y el subgrupo 2B agrupa por tres tequilas más dispersos: TB01, TB32 y TBM34.

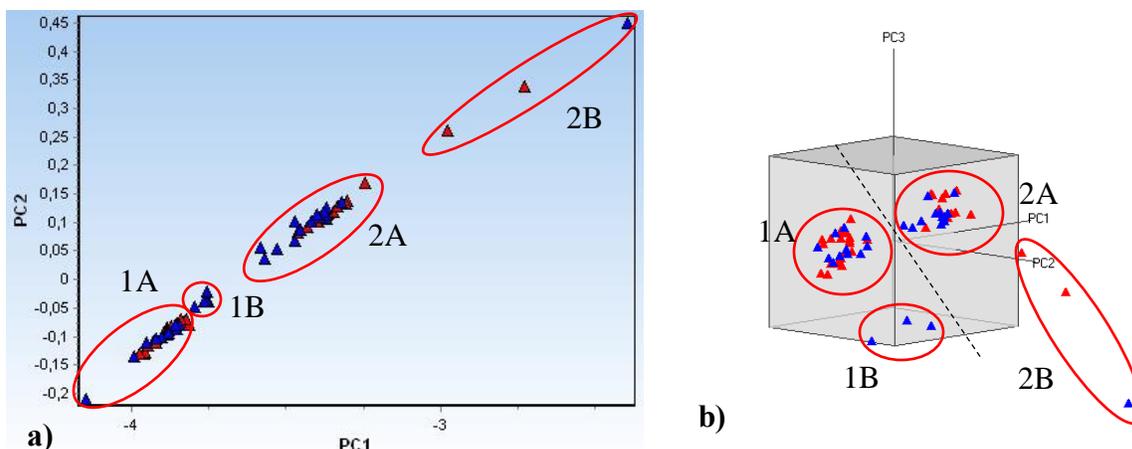


Figura 5. Diagrama de scores exploratorio en intervalo 400 a 4000 cm^{-1} sin escalado.
5a) Diagrama PC1-PC2 5b) Diagrama PC1-PC2-PC3.

La ubicación de los scores de los tequilas en los diferentes grupos de la figura PC1-PC2-PC3 está en función de su contenido alcohólico. Este comportamiento se muestra cualitativamente en figura 2

En la figura 6 en la zona de 3000 a 4000 cm^{-1} , aparece la banda típica de interacción O-H, donde se señalan los subgrupos antes mencionados. De esta forma, puede observarse que el contenido alcohólico va en aumento desde el subgrupo 1A, hasta el subgrupo 2B. Así, el TBM34 es de mayor porcentaje de alcohol.

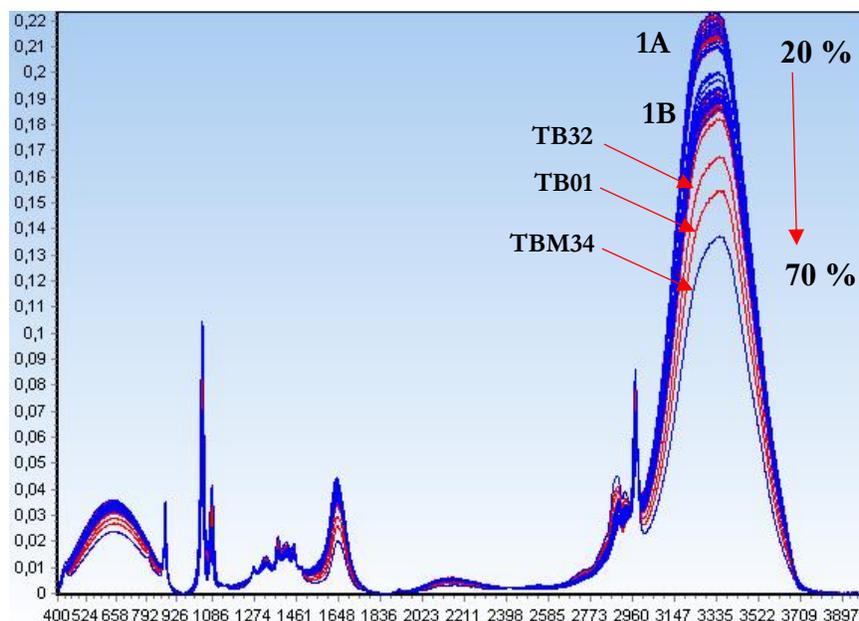


Figura 6. Espectros FTIR de las muestras de estudio en un intervalo de 400–4000 cm^{-1} .

Fueron elaborados dos modelos preliminares más utilizando diferentes escalados (figura 7). La figura 7a muestra el diagrama de escores con mean center, mientras que en 7b se utilizó autoescalado. En estos modelos al igual que en el modelo previo sin escalado (figura 5) observamos la formación de los dos diferentes grupos antes mencionados. De esta manera se puede concluir que PC1 determina principalmente la diferenciación de los tequilas por contenido alcohólico.

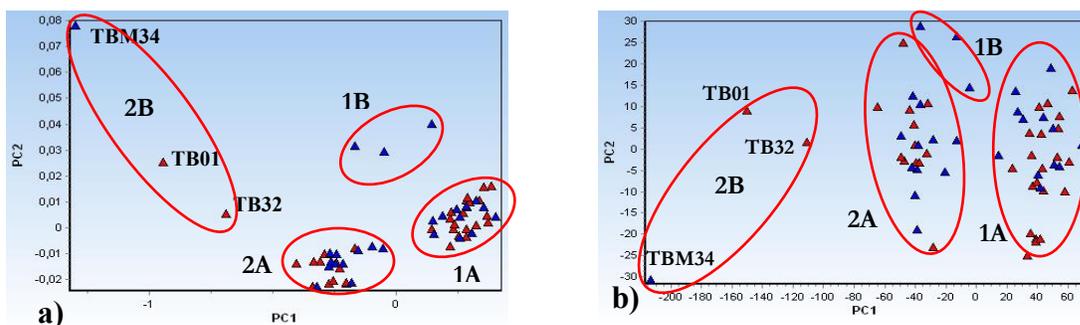


Figura 7. Diagramas de scores en un intervalo de 4000-400 cm^{-1}

7a) Diagrama PC1-PC2 con mean center 7b) Diagrama PC1-PC2 con autoescalado

En la tabla siguiente se muestran los resultados de varianza para cada modelo.

Tabla 8. Porcentajes de varianza explicada para 2 PCs.

Pretratamiento	Varianza	
	PC1	PC2
Mean Center	99.58%	0.24%
Autoescalado	83.56%	4.11%

Para construir los modelos, no se tomó en cuenta el tequila TBM34 ya que su contenido alcohólico se encuentra en 79% y el tequila TBM03 ya que su contenido alcohólico está en 20%.

En la siguiente sección se muestran los modelos obtenidos con diferentes intervalos con datos autoescalados, ya que resultó ser el óptimo para la diferenciación de los tequilas por categoría.

Intervalos de trabajo para modelos PCA.

Los modelos presentados a continuación siguen la nomenclatura descrita en la sección 5; color rojo para tequilas 100 % agave y azul para tequilas mixtos.

Para determinar el modelo óptimo se consideraron únicamente tres componentes principales. Fueron descartadas las muestras identificadas previamente como anómalas.

Los siguientes gráficos de scores muestran el resultado de los 4 intervalos usados. El principal criterio para la elección de los intervalos se basó en la elección de zonas con mayor contribución de los principales compuestos del tequila, los cuales fueron descritos en la tabla 7.

Intervalo 1. 400-1550 cm^{-1} .

La figura 8 muestra los diferentes diagramas de scores; tres PCs explican el 96.70% de la varianza total (PC1 es igual a 86.09%, PC2 a 5.91% y PC3 aporta el 4.70%). En la figura 8a se observa el sub-espacio PC2 vs PC3 en el que las muestras presentan una distribución alargada y una clara

formación de grupos con base en la categoría del tequila. El grupo de 100 % agave es el más compacto posiblemente debido a la homogeneidad de su composición química.

La figura 8b presenta el diagrama de scores de PC1 vs PC3, donde se observa una diferencia entre el grupo de los tequilas 100 % agave y mixtos. A lo largo de PC1 se observa la diferenciación por contenido alcohólico, incluso quitando las bandas de 3000 a 4000 cm^{-1} (banda principal del grupo OH). Las muestras TBM73 y TB01 aparecen a valores extremos de PC1 por presentar altos contenidos alcohólicos. A través del eje PC2 se observa la diferenciación por categorías en ambas figuras (a y b).

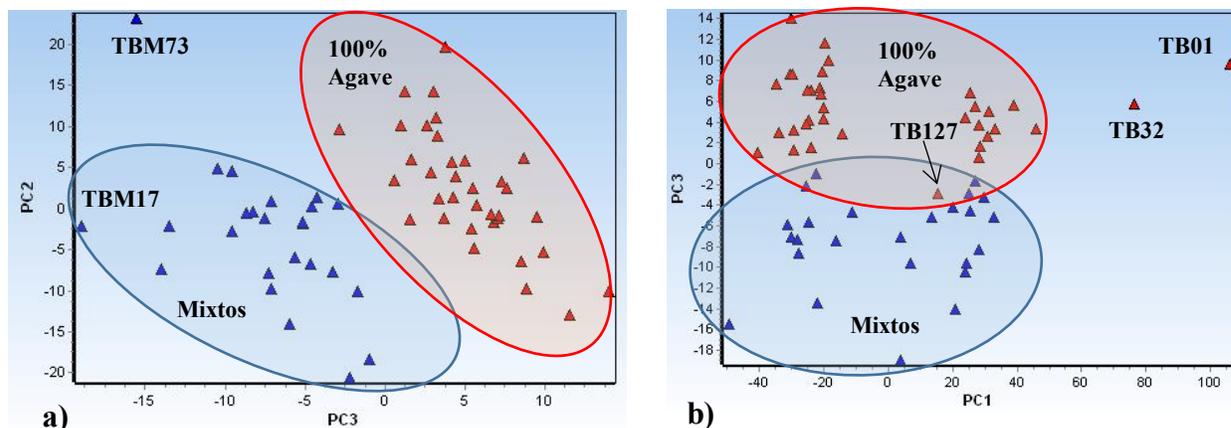


Figura 8. Diagrama de scores autoescalado, intervalo 400 a 1550 cm^{-1}
a) PC2 vs PC3 b) PC1 vs PC3.

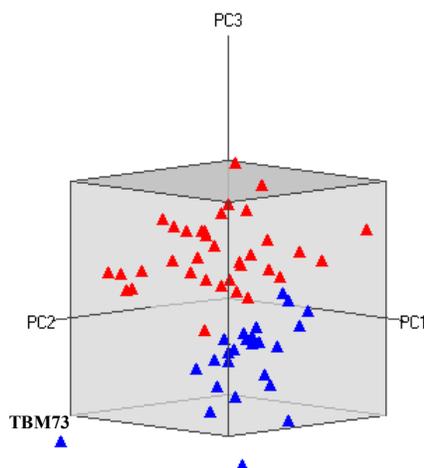


Figura 9. Diagrama de scores en 3 dimensiones (PC1-PC2-PC3) con datos autoescalados en el intervalo de 400 a 1550 cm^{-1} .

En la figura 8b, los tequilas TB01 y TB32 se mantienen dentro de la zona de tequilas 100 % agave, pero el eje PC1 los separa del resto de tequilas del grupo ya que presentan una mayor concentración de alcohol. La figura 10 muestra la comparación de estos tequilas con un tequila típico 100% agave (TB37).

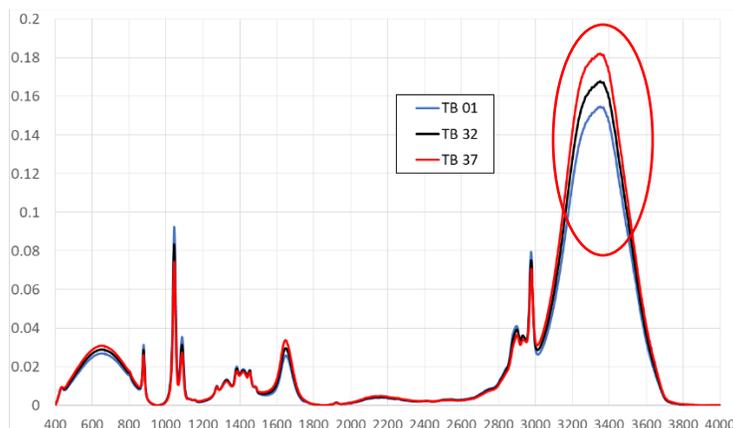


Figura 10. Comparación espectral de Tequilas TB01 (Azul) y TB32 (Negro) con el Tequila TB37 (Rojo).

Las figuras 8a y 9 muestran que el tequila mixto TBM73 que no se agrupa con el resto. Este tequila según información del CRT se elaboró con Frudex al 49%. La figura 11 muestra un acercamiento del intervalo de trabajo de 400 a 1550 cm^{-1} en el cual realizamos la comparación del tequila TBM73 con tequilas típicas de cada categoría; es observable que el tequila TBM73 muestra considerablemente menor absorbancia en sus bandas en todo el intervalo que el resto de los tequilas, principalmente en la zona de 1000 a 1100 cm^{-1} . En esta zona se tienen señales de alcoholes de cadena larga como butanol e isobutanol de los cuales su concentración depende de la cantidad de agave utilizado en la fermentación (37); por lo que al elaborar este tequila con Frudex, es de esperarse que su concentración en estos alcoholes baje y por lo tanto se aleje del resto de tequilas del grupo. Para comprobar esto sería requerido cuantificar estos alcoholes en las muestras de tequila, lo cual se sale del alcance de este trabajo por lo cual nos limitamos a corroborar esto con los trabajos y literatura disponible acerca del tema hasta el día de hoy y tomaremos este hecho en cuenta para futuros trabajos.

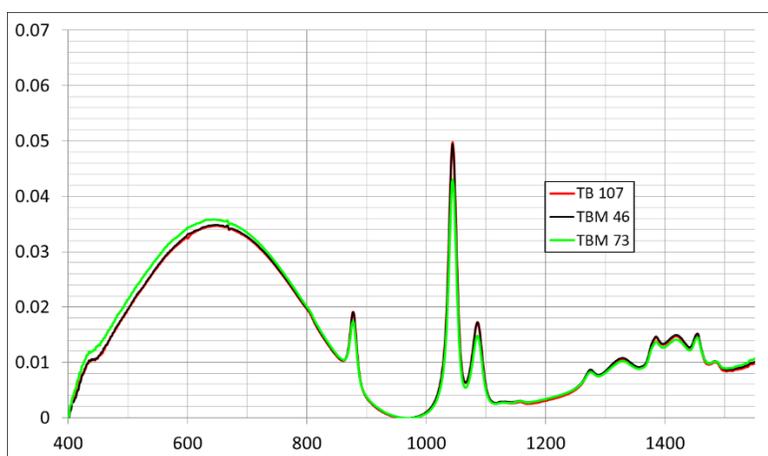


Figura 11. Comparación espectral de Tequila TBM73 (Verde) con dos Tequilas típicos; TB107 (Rojo) y TBM46 (Negro).

Por tanto, podemos concluir que el modelo óptimo en este intervalo es el de la figura 8^a, es decir, autoescalado, intervalo 400 a 1550 cm^{-1} .

Intervalo 2. 850-1550 cm⁻¹.

En este caso tres PCs explican el 97.84% de la varianza total; (PC1= 86.11 %, PC2=8.37 % y PC3=3.36 %). En este intervalo observamos una participación más activa por parte de PC2 para la diferenciación por categoría. En el primer gráfico se tiene el subespacio PC2 vs PC3 en el que se aprecia una contribución de ambos componentes para la diferenciación por categoría; los tequilas TBM73, TBM17 y TBM23 aparecen fuera del grupo de tequilas mixtos debido a sus concentraciones variables de alcohol como se explicó en las figuras 6 y 11. En la figura 12, en el diagrama PC1 vs PC2 se observa la separación de las muestras por categoría a lo largo del eje PC2 con una distribución más horizontal en el eje de las abscisas de los tequilas 100 % agave. Al igual que en el modelo anterior observamos la separación del tequila TBM73 por contenido alcohólico; también, a mayores valores de PC1 observamos a los tequilas TB01 y TB32 (tequilas con mayor concentración de alcohol del grupo).

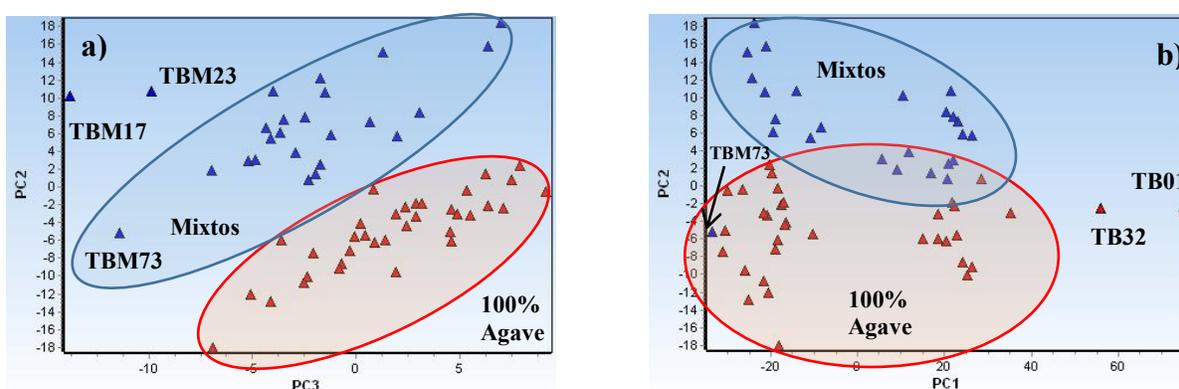


Figura 12. Diagrama de scores autoescalado, intervalo 850 a 1550 cm⁻¹.
a) PC2 vs PC3, b) PC1 vs PC2

El diagrama en tres dimensiones (figura 13) muestra la contribución de los tres PCs para la diferenciación por categoría donde PC2 y PC3 son los que determinan principalmente este acomodo, mientras que PC1 gasta la mayoría de su varianza en la diferenciación por contenido alcohólico y una pequeña parte contribuye a la diferenciación por categoría.

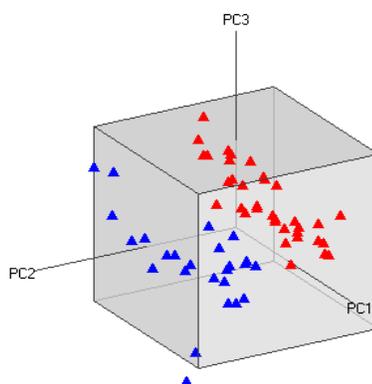


Figura 13. Diagrama de scores PC1-PC2-PC3, autoescalado, intervalo de 850 – 1550 cm⁻¹.

Intervalo 3. 400 a 1800 cm^{-1} .

Para el tercer modelo (figura 14) se amplió el intervalo de número de onda hasta 1800 cm^{-1} , a fin de incorporar la señal del grupo carbonilo (1500 a 1800 cm^{-1}); sin embargo, estas señales están enmascaradas por la absorción del agua, por lo que al incluirlas, no se advierte una mejora en el modelo. La separación por categoría de los tequilas se observa a lo largo del eje PC3 y en el caso de los tequilas 100 % agave, se forman dos subgrupos en función de su contenido alcohólico. A mayor valor de PC1, mayor es este contenido. Tres PCs explican el 97 % de la varianza total (PC1=87.9 0%, PC2=5.04 %, PC3=4.02 %).

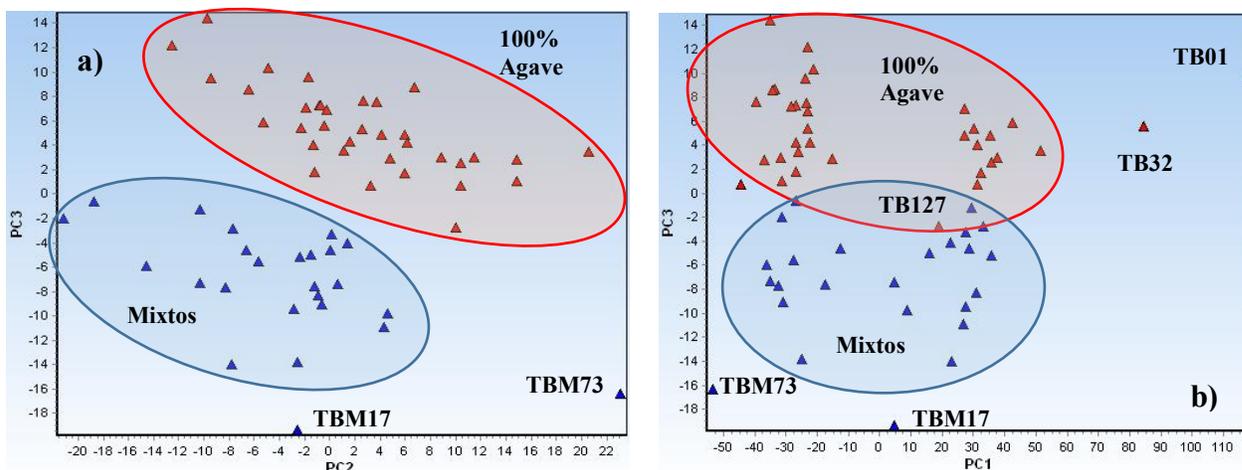


Figura 14. Diagrama de scores autoescalado, intervalo 400 - 1800 cm^{-1} .
a) PC2 vs PC3 b) PC1 vs PC3

En el sub-espacio PC2 vs PC3, al descartar el eje de PC1, ya no se observa el solapamiento entre las categorías de tequilas.

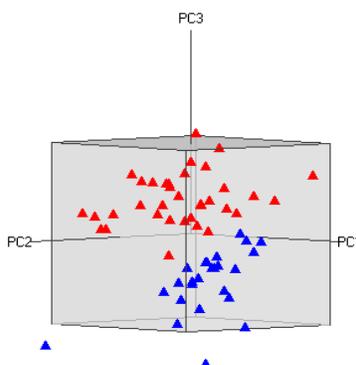


Figura 15. Diagrama de scores PC1-PC2-PC3, autoescalados, intervalo 850 - 1550 cm^{-1} .

Intervalo 4. 850 a 1800 cm^{-1}

Para este último intervalo se consideraron las señales comprendidas de 850 cm^{-1} a 1800 cm^{-1} . En este caso se quitan señales características de alcoholes como el metanol, etanol, propanol e isopropanol. No obstante, las señales principales de algunos otros alcoholes como el isobutanol y el butanol (que tienen que ver con la diferenciación por categoría), se mantienen dentro de este

intervalo y vibraciones de estiramiento del enlace C-OH (1075 a 1025 cm^{-1}). Tres PCs explican el 97.93 % de la varianza acumulada, donde $PC1=88.88\%$, $PC2=6.49\%$, $PC3=2.56\%$.

Sin embargo, este último modelo no presenta gran diferencia con los previos.

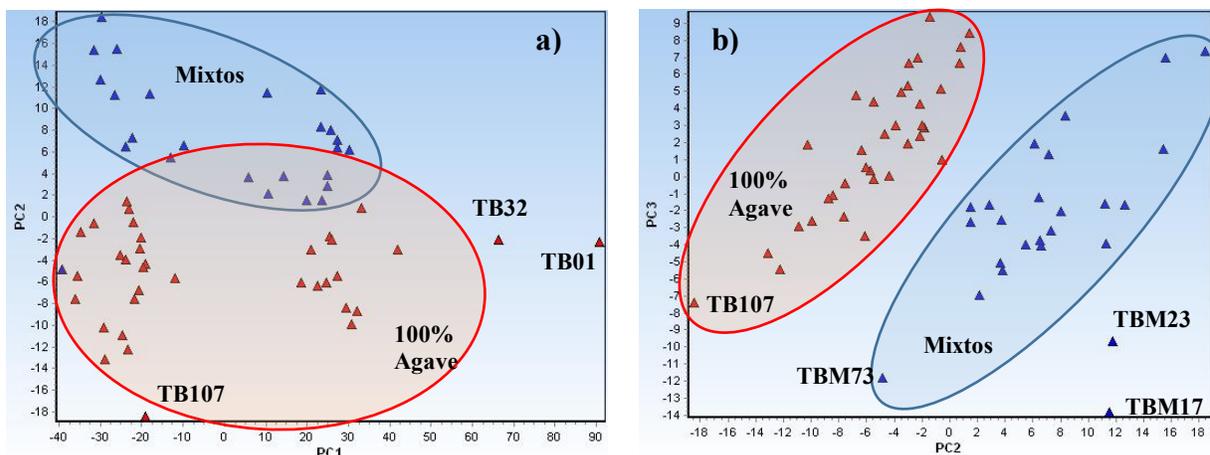


Figura 16. Diagrama de scores a) PC1 vs PC2, b) PC2 vs PC3 autoescalado, 850 a 1800 cm^{-1} .

En los diagramas de scores anteriores se observa la separación del tequila TB107 el cual aparece a ligeramente separado del resto de tequilas del grupo, esta separación es debido a

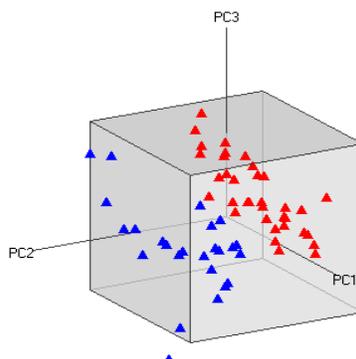


Figura 17. Diagrama de scores en 3 dimensiones con datos autoescalados en el intervalo de 850 a 1800 cm^{-1} (PC1-PC2-PC3).

La tabla siguiente (tabla 9), muestra un resumen de los valores de varianza obtenidos para cada PC en cada uno de los casos.

Tabla 9. Porcentajes de varianza explicada de 3PCs para cada modelo.

Modelo	Intervalo número de onda (cm^{-1})	PCs			Varianza acumulada
		PC1	PC2	PC3	
1	400-1550	86.09%	5.91%	4.70%	96.70%
2	850-1550	86.11%	8.37%	3.36%	97.84%
3	400-1800	87.90%	5.04%	4.02%	97.00%
4	850-1800	88.88%	6.49%	2.56%	97.93%

En todos los modelos prácticamente los modelos son equivalentes al igual que las varianzas (tabla 9). Sin embargo, el subespacio PC2-PC3 es que muestra una diferencia más clara en donde es el PC3 el que determina principalmente esta diferencia. Sin embargo, se podría optar por el intervalo de 400 a 1800 cm^{-1} (modelo 3) donde el diagrama 14 a) guarda una separación de ambos grupos más horizontal a lo largo de PC3.

8.2.2 Análisis de Loadings

Por fines prácticos, fueron calculados los loadings de PC2 y PC3 del modelo óptimo, construido en el intervalo de 400 a 1800 cm^{-1} , ya que ambos PC que determinan la separación por categoría.

Las principales señales que observamos en ambos loadings coinciden con las señales de los principales compuestos en los espectros del tequila dentro del intervalo del espectro seleccionado. Esta zona contiene las señales de los principales compuestos obtenidos durante la fermentación, sin tomar en cuenta los alcoholes que en esta zona muestran algunas señales complementarias.

Para analizar grupos funcionales asignados a los PCs que influyen en la diferenciación de tequilas 100 % agave y mixtos, se asignaron y componentes a cada señal observada en los loadings.

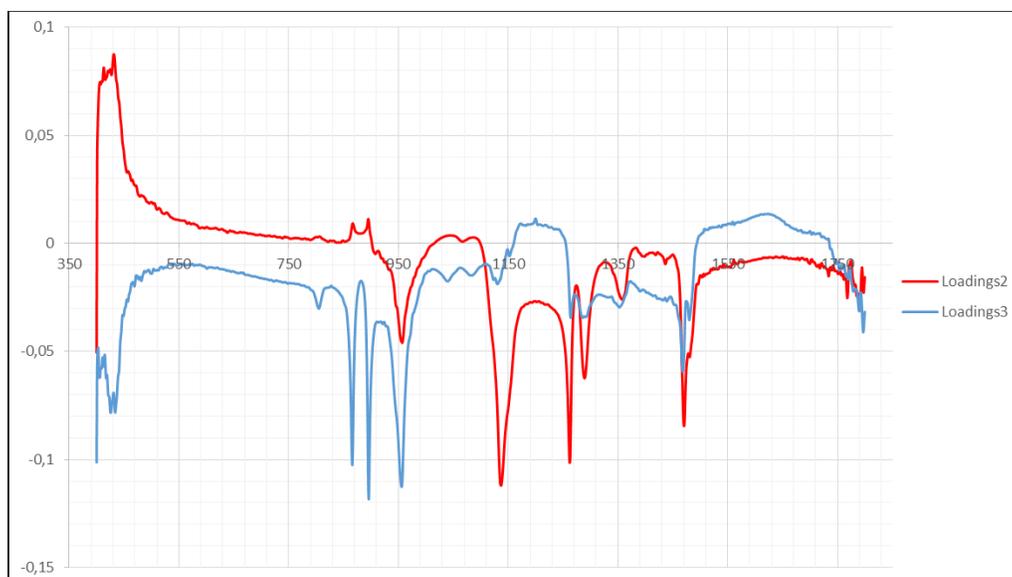


Figura 18. Gráfico de Loadings para modelo obtenido en el intervalo de 400 a 1800 cm^{-1} .

La tabla 10 resume la asignación a los posibles compuestos asignados a cada loading de cada componente según el grupo funcional, de acuerdo con la composición del tequila.

Tabla 10. Compuestos principales del tequila y su asignación de los compuestos a los loadings de PC2 y PC3 de 1800 a 400 cm^{-1} .

Variable	Componente		Compuestos
1550			2-metil furil; 2,5-dimetil furano; propanol; 1-butanol, 3-metil acetato
1500	PC2	PC3	5-metil furfural; Furfural; benzaldehído; 3-metil butanol; ácido acético, etil éster; acetaldehído; 2-metil furil; 2,5-dimetil furano; isobutanol; 1-butanol, 3-metil acetato; propanol
1450	PC2	PC3	Furfural, benzaldehído; 3-metil butanol; ácido acético, etil éster; acetaldehído; 2-metil furil; 2,5-dimetil furano; isobutanol; 1-butanol, 3-metil acetato
1350	PC2	PC3	3-metil butanol; etil éster; acetaldehído; 2-metil furil; 2,5-dimetil furano; acetato 2-furanmetanol; isobutanol; 1-butanol, 3-metil acetato; oxalato
1300	PC2	PC3	Benzaldehído; fenil etanol; ácido acético; 2-metil furil; isobutanol; 2-fenil etil acetato; 1-butanol, 3-metil acetato
1250	PC2	PC3	Furfural, metanol; ácido acético; etil éster; 2,5-dimetil furano
1150	PC2	PC3	Furfural; fenil etanol; acetaldehído; 2-metil furil
1080		PC3	Furfural; 3-metil butanol; etil éster; acetaldehído; 2-metil furil; ácido butanóico
1050	PC2	PC3	Furfural; 3-metil butanol; etil éster; 2,5-dimetil furano; propanol; isobutanol; 1-butanol, 3-metil acetato
1040		PC3	5-metil furfural; 2-metil furil; 2,5-dimetil furano; isobutanol; 2-fenil etil acetato; 1-butanol, 3-metil acetato; ácido butanóico; alfa-terpineol
950	PC2	PC3	Furfural; fenil etanol; ácido acético; etil éster; 2,5-dimetil furano; propanol; 2-fenil etil acetato
900	PC2	PC3	Furfural; fenil etanol; 3-metil butanol; alfa-terpineol; ácido acético; etil éster; acetaldehído; 2-metil furil; 2,5-dimetil furano, linalol; propanol, isobutanol; 4-terpineol; oxalato
850	PC2	PC3	Etil éster; acetaldehído
800		PC3	Furfural; Benzaldehído; metanol; alfa terpineol; 2-metil furil; 2,5-dimetil furano; etanol; oxalato
450	PC2	PC3	Benzaldehído
400	PC2	PC3	Benzaldehído; metanol; etil éster

En la tabla 10 podemos observar tanto PC2 como PC3 están definidos entorno a variables muy semejantes. Sin embargo, dado que los componentes principales son ortogonales por definición, PC2 y PC3 deben venir definidos por variables diferentes, es decir, que ambos factores están teniendo en cuenta compuestos de estructura química diferente como se puede observar en la tabla.

Cabe hacer notar que las variables 1080, 1040 y 800 cm^{-1} sólo dan señales para PC3 y éstas están vinculadas a compuestos furánicos entre otros. Con relación a la variable de 1500 cm^{-1} , su señal presenta una mayor intensidad en PC2 que en PC3; esta variable relacionada con derivados furánicos (furfural; 2-metil furil y 2,5-dimetil furano). Las variables entorno a la zona de 1080 a 1040 cm^{-1} del loading de PC3 se pueden asociar a los compuestos con grupos hidroxilo como el isobutanol, 3-metil butanol y el 1-butanol, 3-metil acetato. La mayoría de las señales en esta zona se deben a estiramientos del enlace CH₂-OH, posiblemente de alcoholes de cadena larga, cuyo contenido depende de la fuente de azúcares utilizados para la fermentación.

En 1450 cm^{-1} tenemos señales muy pequeñas en PC2 que pueden ser asignadas a compuestos como 2,5-dimetil furano y etil éster en los cuales los grupos CH₃ y CH₂ generan señales de flexiones simétricas y asimétricas.

La variable en torno a 1350 cm^{-1} se relaciona con grupos como el isobutanol, 2,5-dimetil furano, el pentanol y el oxalato; en el caso de los alcoholes se puede explicar que esta variable aporte importancia a la diferenciación por las pectinas disponibles para la fermentación con el uso de

agave y éstas suelen ser más en los tequilas 100% agave y generan mayor concentración de estos alcoholes (37). Por otro lado, el oxalato, como se ya mencionó, se encuentra en mayor medida en la planta de agave.

Las variables en 1300 y 1250 pueden ser originadas por los grupos fenil etanol, ácido acético, etil éster, 2,5-dimetil furano. Estas señales son generadas principalmente por estiramientos del enlace C-O.

La variable 1150 es muy similar en ambos loadings incluso en intensidad y esta es asignable a grupos como el acetaldehído y el fenil etanol.

La variable 950 cm^{-1} difiere un poco en la forma de la banda entre ambos loadings; sin embargo, puede ser asignada a grupos como el ácido acético y el 2,5-dimetil furano. Un caso similar es variable de 900 cm^{-1} la cual también es asignable a estos compuestos, pero también se tiene la contribución de otros compuestos furánicos como el furfural y el 2-metil furil y no furánicos como el oxalato.

Por último, las señales observadas en 850 a 400 cm^{-1} son más específicas y con base en la tabla 10 sólo pueden corresponder a compuestos como el etil éster, acetaldehído, benzaldehído y compuestos furánicos. Estas variables junto las incluidas en la zona de 1080 y 1040 cm^{-1} son de las más importantes en PC3 y por lo tanto las que más aportan a la diferenciación entre categorías, ya que son muy específicas en su forma y las de mayor intensidad a las observadas en el loading de PC2.

8.2.3 Análisis Cluster (CA)

A partir de las condiciones establecidas para los modelos PCA, se llevó a cabo el análisis cluster. Dos intervalos fueron los que mostraron mejor resultado para la clasificación de los tequilas por categoría.

La primera exploración se realizó a partir de datos brutos con Autoescalado. No obstante, la separación resultante fue por contenido alcohólico y no por categoría. Debido a esto, los modelos se construyeron considerando una matriz de 10 scores desde PC2 a PC11 obtenidos en los modelos PCA en cada intervalo específico; no se utilizó PC1 para evitar la influencia por contenido etanólico. A continuación, se presentan los modelos en cada intervalo.

Intervalo 1. 400 a 1550 cm^{-1}

El dendrograma de la figura 19 corresponde al intervalo de 400 a 1550 cm^{-1} utilizando método Ward y como distancia, la correlación de Pearson, la cual muestra la separación de los tequilas por categoría; a la izquierda observamos los tequilas 100 % agave (grupo A) y del lado derecho los tequilas mixtos (grupo B).

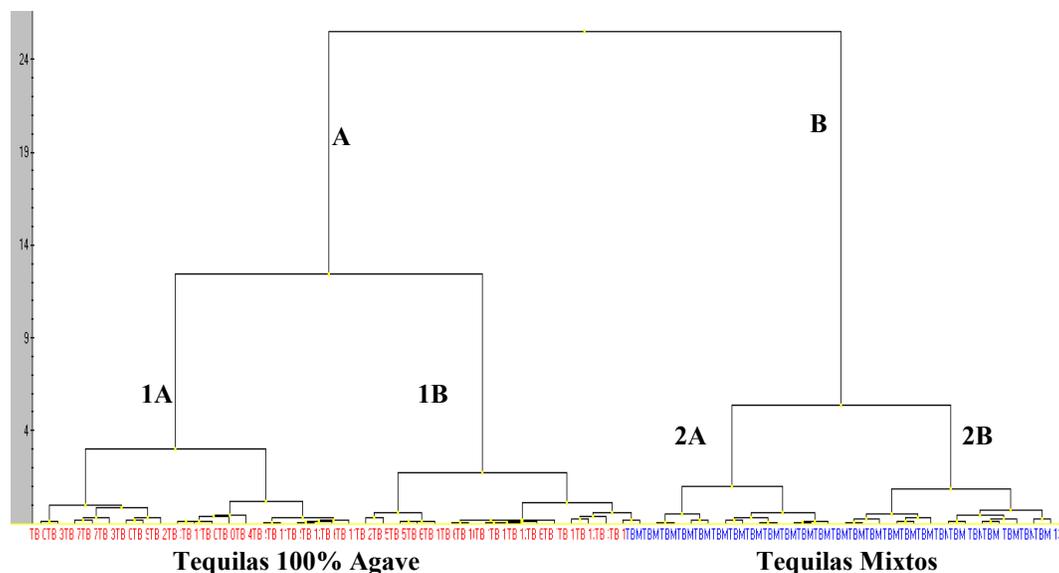


Figura 19. Dendrograma obtenido con método Ward y correlación de Pearson como medida de similitud en el intervalo de 400 a 1500 cm^{-1} .

Cada grupo se divide en dos subgrupos A y B; los subgrupos 2A y 2B contienen a todos los tequilas mixtos exceptuando uno. El subgrupo 1A contiene sólo tequilas 100 % agave, en tanto que el 1B contiene al resto y un tequila mixto.

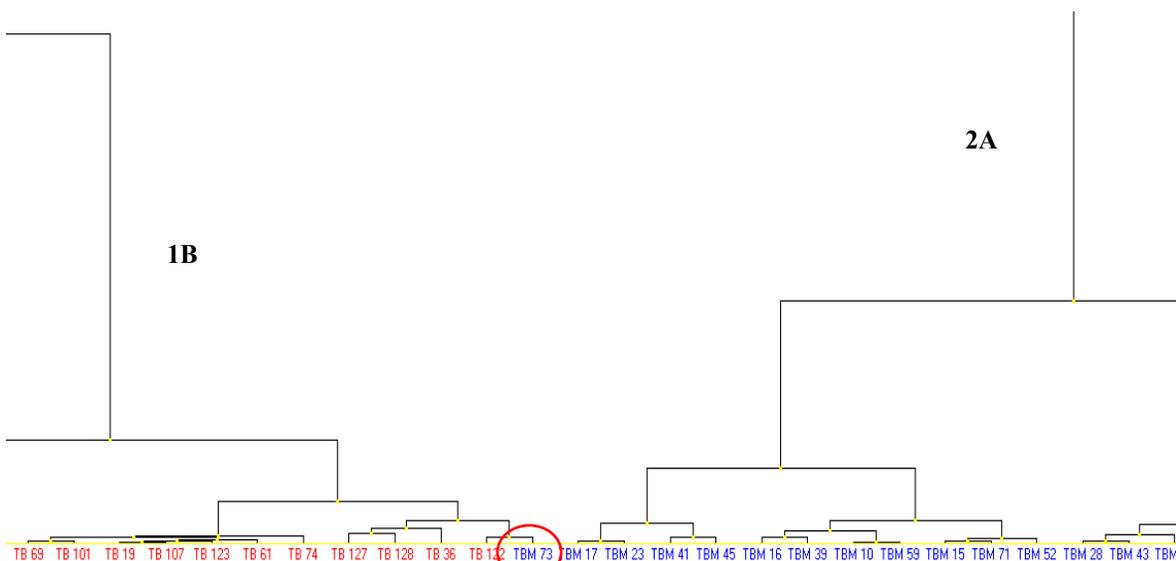


Figura 20. Acercamiento de subgrupos observados en la figura 19.

Un acercamiento de este subgrupo 1B permite identificar a este tequila como el TBM73 (Figura 20). Este tequila también se diferenciaba en PCA del resto de los mixtos por estar enriquecido con Frudex, además de que pudiera tener una composición parecida a los 100 % agave.

Intervalo 2. 850 a 1550 cm^{-1}

En este caso se quitó la zona de 400 a 849 cm^{-1} (señales de alcoholes y algunas de compuestos aromáticos). La Figura 21 muestra el dendrograma en el intervalo de 850 a 1550 cm^{-1} , obtenido

con Ward como método y correlación de Pearson como medida de similitud. Una formación de dos grupos reúne a todos los tequilas, observando una clara separación por categoría.

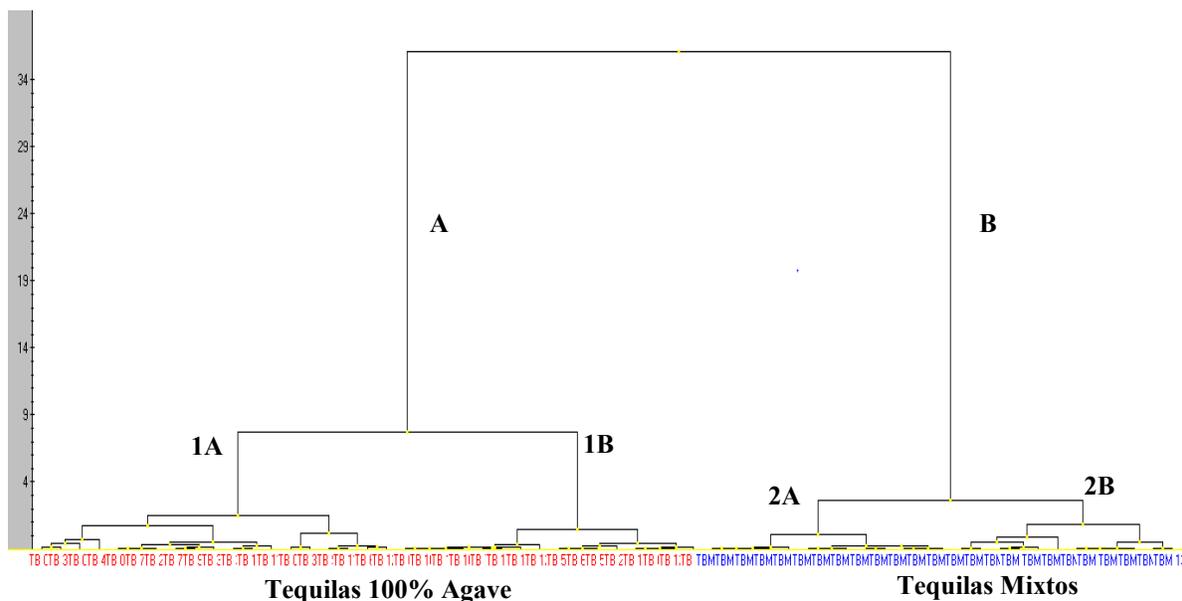


Figura 21. Dendrograma obtenido con método Ward y correlación de Pearson como medida de similitud en el intervalo de 850 a 1500 cm^{-1} .

El conjunto A contiene todos los tequilas 100 % agave y en el B los mixtos; al evitar la zona de 400 a 849 cm^{-1} , el tequila TBM73 se agrupó con el resto de los tequilas de su grupo. Por tanto, se puede concluir que este modelo es el óptimo.

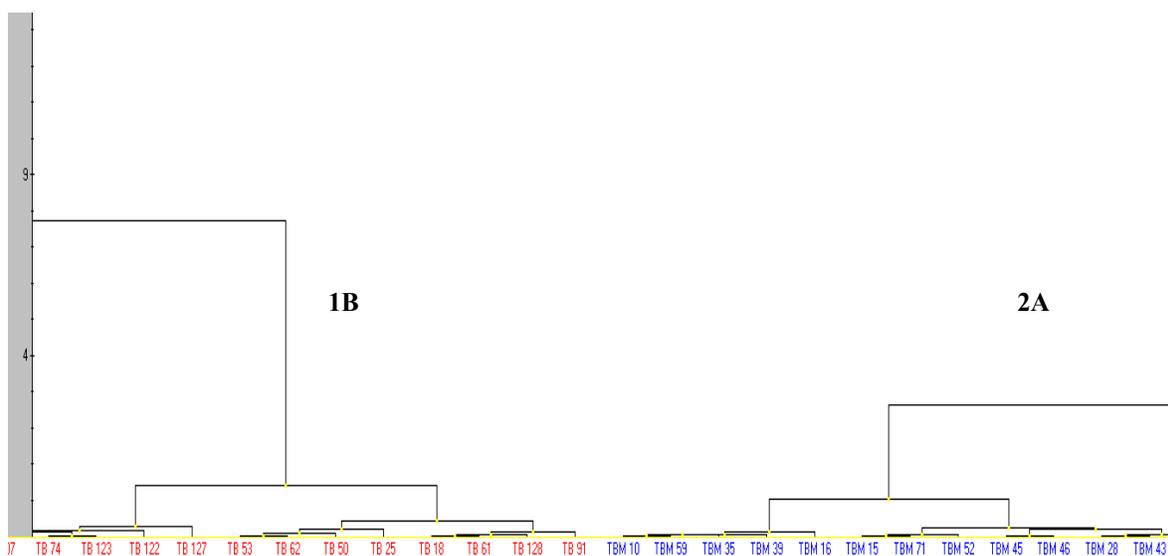


Figura 22. Acercamiento de subgrupos observados en la figura 21.

8.3 Modelos obtenidos bajo técnicas supervisadas.

8.3.1 Curvas de Potencia (PC).

Para la construcción de los modelos por Curvas de Potencia, la matriz con 10 Scores (scores de PC2 a PC11), ya que la matriz original en los intervalos propuestos no dio buenos resultados. Al igual que los modelos anteriores de CA, fueron utilizados los mismos intervalos de trabajo.

Las elipses de iso-probabilidad obtenidas para cada intervalo son mostradas en las figuras 23 y 24; dichas elipses se construyeron partiendo de un conjunto de calibración que incluye todas las muestras medidas anteriormente, exceptuando las anómalas y un segundo conjunto compuesto por 4 muestras de cada categoría denominado “muestras test” o “conjunto de validación”. Las muestras test fueron elegidas de manera que fueran representativas de cada categoría.

Las 4 muestras test para los tequilas 100 % agave fueron: TB91, TB92, TB101 y TB127 y para los tequilas mixtos: TBM10, TBM41, TBM97 y TBM131. Los resultados de predicción para cada muestra se observan en las tablas para cada intervalo (tablas 11 y 12).

Intervalo 1. 400 a 1550 cm^{-1}

La figura 23 muestra dos conjuntos de las elipses de iso-probabilidad en el intervalo de 400-1550 cm^{-1} correspondientes a cada categoría con un leve traslape en la frontera de cada conjunto donde se ubica la muestra test TB127. Esta separación del TB127 puede deberse a su contenido alcohólico como se observó en los modelos PCA.

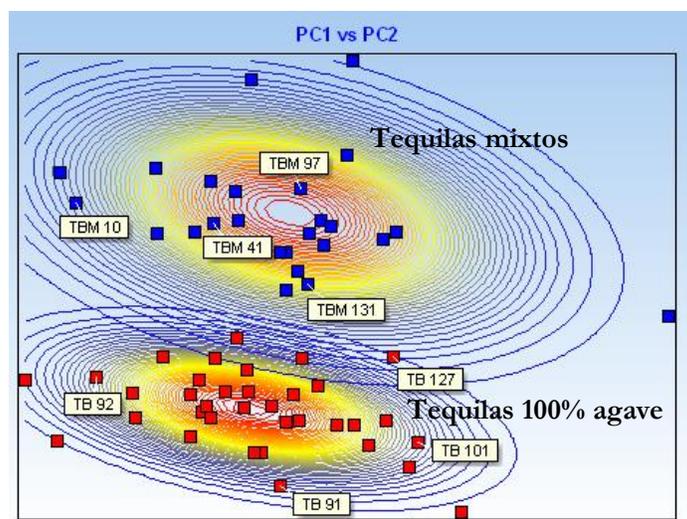


Figura 23. Elipses de probabilidad obtenidas con matriz de scores en el intervalo de 400 a 1550 cm^{-1} .

Con base en los resultados de probabilidad (tabla 11) podemos decir que todas las muestras de validación fueron asignadas correctamente, excepto la muestra TB127 que presenta porcentajes de pertenencia similares para ambas categorías. Los tequilas 100 % agave presentan un agrupamiento compacto debido a que sus matrices son más similares, a diferencia de la elipse de los tequilas mixtos, que al presentar una mayor dispersión, permite que más muestras test queden en el centro, aumentando su porcentaje de pertenencia al grupo.

Tabla 11. Porcentaje de pertenencia a cada categoría para las muestras test.

Muestra	PC1	PC2	Porcentaje de pertenencia	
			Tequila 100%	Tequila Mixto
TB91	5.3960	-9.2120	9.87	0.00
TB92	-10.4158	-7.9000	23.35	0.00
TB101	13.8468	-2.5675	21.62	0.00
TB127	9.8029	1.9033	1.68	4.44
TBM10	-17.8184	2.6979	0.00	16.12
TBM41	-6.4510	5.2214	0.00	76.76
TBM97	-1.8513	9.7382	0.00	91.39
TBM131	1.8932	4.0004	0.00	43.46

Los tequilas mixtos presentan mayor porcentaje de pertenencia al grupo, llegando incluso a valores por arriba del 90 % (caso del tequila TBM97).

Intervalo 2. 850 a 1550 cm^{-1}

Las elipses resultantes del segundo intervalo, de 850 a 1550 cm^{-1} , se muestran en la figura 24. Para este caso observamos un agrupamiento similar al observado en el modelo anterior, pero con la elipse de tequilas 100% agave un tanto más compacta. Por otro lado, la mayoría de las muestras test de los tequilas mixtos se ubican más cerca del centro de la elipse. Nuevamente existe un leve traslape entre las fronteras de ambos grupos.

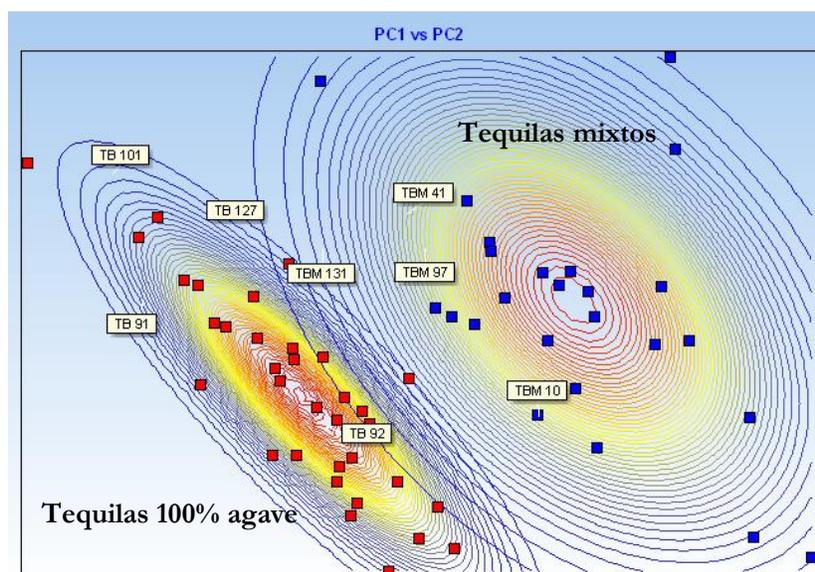


Figura 24. Elipses de probabilidad obtenidas con matriz de scores en el intervalo de 850 a 1550 cm^{-1} .

La Tabla 12 muestra los porcentajes de pertenencia resultantes de este modelo. Para este caso, sólo el porcentaje de pertenencia para 2 tequilas mixtos está por arriba del 90 %; en cambio, en tequilas 100 % agave estos porcentajes son bajos. La muestra TB127 que presenta porcentajes de pertenencia similares para ambas categorías

Tabla 12. Porcentaje de pertenencia a cada categoría para las muestras test.

Muestra	PC1	PC2	Porcentaje de pertenencia	
			Tequila 100%	Tequila Mixto
TB91	-9.6916	-1.1115	11.04	0.00
TB92	1.7090	-8.1090	22.86	0.00
TB101	-11.5785	6.0860	14.32	0.00
TB127	-5.6725	4.0596	3.03	3.04
TBM10	15.1718	-7.6420	0.00	9.87
TBM41	6.9444	3.7556	0.00	96.58
TBM97	8.0501	1.7860	0.00	98.07
TBM131	1.6176	1.7943	0.00	44.43

Con base en estos resultados, se puede concluir que pese a que se tiene una satisfactoria separación de las elipses de Iso-probabilidad en los gráficos, el modelo que la asignación de las muestras test a cada categoría no es satisfactoria.

8.3.2 Funciones de Potencia (PF)

A partir de esta técnica, los modelos fueron construidos con el programa Matlab; para esto primero se optimizaron los parámetros como el valor del Kernel y otros particulares de cada técnica. Los resultados mostrados son los obtenidos con la matriz original en los intervalos propuestos anteriormente (400 a 1550 y 850 a 1550 cm^{-1}). El pretratamiento elegido fue Autoescalado, pero éste se realizó en el programa Matlab.

Intervalo 1. 400 a 1550 cm^{-1} .

Los parámetros elegidos para determinar los valores de sensibilidad, especificidad y suavizado para el primer modelo construido en el intervalo de 1550 a 400 cm^{-1} fueron los siguientes:

- Pretratamiento: Autoescalado.
- Kernel: Gaussiano.
- Percentil por clase: 95
- Tipo de validación: Validación cruzada (persianas venecianas)
- Número de CV: 10

La figura 25 muestra la tasa de error para cada suavizado por clase, observando que la sensibilidad y especificidad aumentan con la disminución de la tasa de error por lo que el modelo a estos valores de suavizado hace una clasificación más certera para las muestras test. La clase 1 representa a los tequilas 100 % agave y la clase 2 a los tequilas mixtos.

Es importante hacer notar que para la obtención de estos parámetros se utilizó la matriz de calibración la cual contiene todas las muestras de estudio salvo las muestras test, las cuales se especificaron en los modelos obtenidos por la técnica de curvas de potencia.

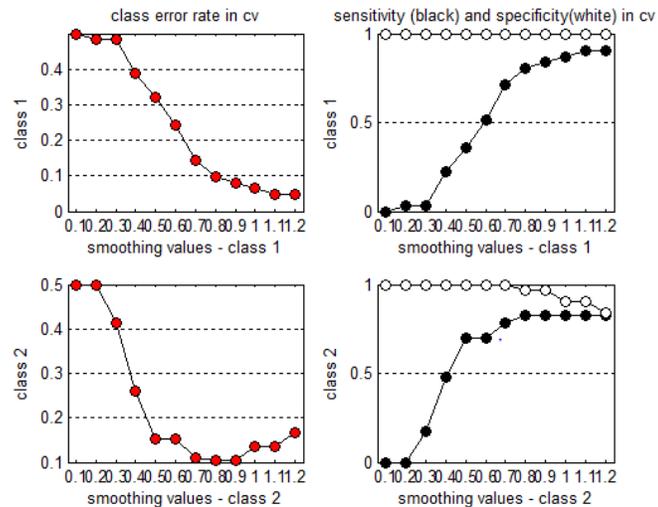


Figura 25. Tasa de error vs suavizado; sensibilidad y especificidad vs suavizado para cada clase.

Con los resultados de la figura 25 se determinó el suavizado para la clase 1 en 1.2 y el suavizado para la clase 2 en 0.80. Con estos valores y los parámetros antes mencionados, el modelo se corrió con los siguientes criterios:

- Pretratamiento: Autoescalado.
- Kernel: Gaussiano.
- Percentil por clase: 95
- Tipo de validación: Validación cruzada (persianas venecianas)
- Número de CV: 10
- Suavizado Clase 1: 1.2
- Suavizado Clase 2: 0.8

Una vez calculado el modelo se realizó la predicción de las muestras test; la figura 27 muestra las funciones de potencia para cada clase con un buen agrupamiento de las muestras test con su respectiva clase. La tasa de error obtenida en el modelo de calibración fue de 0, exactitud de 1 y un 0.09 de muestras no asignadas (Consultar anexos).

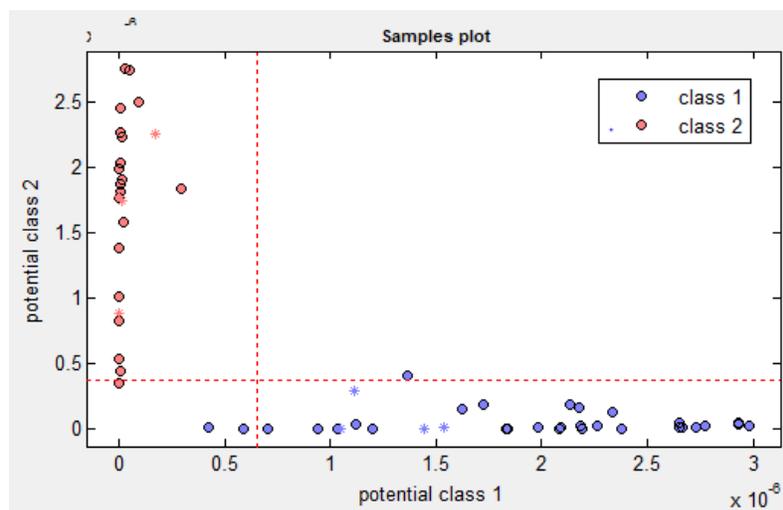


Figura 26. Función de Potencia Clase 1 vs Función de Potencia Clase 2.

Para las muestras test la tasa de error fue de 0, se obtuvo una exactitud de 1 y las muestras no asignadas 0.15.

En la tabla 13 se tienen los valores obtenidos de especificidad, sensibilidad y precisión para el modelo tanto de calibración (fitting) y el de validación. Para cada caso y clase el valor obtenido fue de 1, lo que indica que el modelo es óptimo para llevar a cabo la clasificación de las muestras por categoría.

Tabla 13. Valores de Especificidad, Sensibilidad y Precisión para modelo de calibración y validación en el intervalo de 400 a 1550 cm^{-1} .

Modelo	Clase	Especificidad	Sensibilidad	Precisión
Fitting	1	1	1	1
	2	1	1	1
Validación	1	1	1	1
	2	1	1	1

Tabla 14. Matriz de confusión para el modelo construido en el intervalo de 400 a 1550 cm^{-1} .

	Clase 1	Clase 2	No asignados
Clase 1	4	0	0
Clase 2	0	4	0

La matriz de confusión que se muestra en la tabla 13 indica que todas las muestras test han sido clasificadas y todas asignadas a la clase correcta (4 por cada clase).

Intervalo 2. 850 a 1550 cm^{-1} .

El cálculo del segundo modelo se realiza en el intervalo de 850 a 1550 cm^{-1} ; siguiendo el mismo procedimiento anterior, primero se obtuvo los valores óptimos de suavizado; los criterios considerados para el cálculo fueron los siguientes:

- Pretratamiento: Autoescalado.
- Kernel: Gaussiano.
- Percentil por clase: 95
- Tipo de validación: Validación cruzada (persianas venecianas)
- Número de CV: 10

En la figura 27 se presentan las curvas de tasa de error, sensibilidad y especificidad vs suavizado muestran los resultados del modelo.

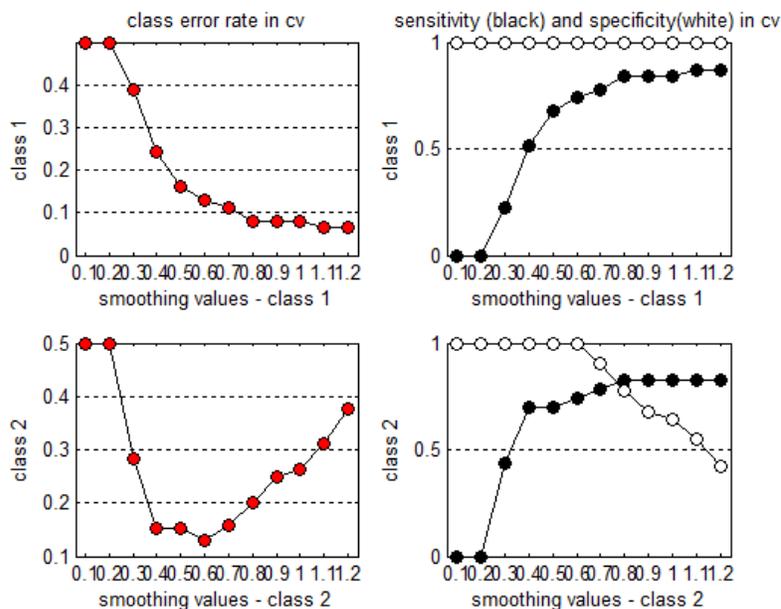


Figura 27. Tasa de error vs suavizado; sensibilidad y especificidad vs suavizado para cada clase.

Se determinó como el suavizado óptimo para cada clase en este intervalo como sigue: suavizado para la clase 1 en 1.2 y para la clase 2 en 0.60. Con estos valores el modelo se corrió con los siguientes criterios:

- Pretratamiento: Autoescalado.
- Kernel: Gaussiano.
- Percentil por clase: 95
- Tipo de validación: Validación cruzada (persianas venecianas)
- Número de CV: 10
- Suavizado Clase 1: 1.2
- Suavizado Clase 2: 0.6

La figura 28 expone el resultado de clasificación de las muestras test en el modelo de calibración; todas las muestras fueron correctamente asignadas y se encuentran dentro del conjunto de calibración. La tasa de error obtenida en el modelo de calibración fue de 0, exactitud de 1 y un 0.17 de muestras no asignadas, 0.08 más alto que el modelo anterior (Consultar anexos).

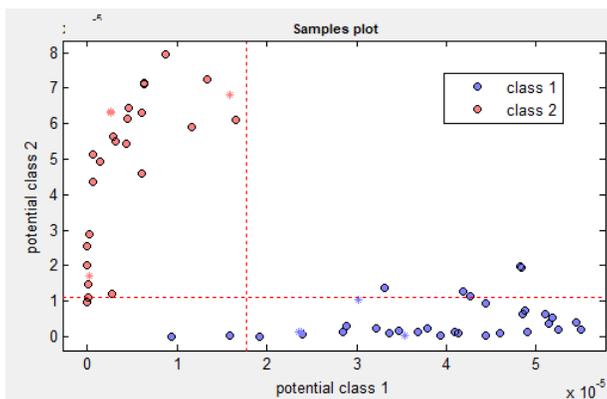


Figura 28. Función de Potencia Clase 1 vs Función de Potencia Clase 2.

Para las muestras test la tasa de error fue de 0, se obtuvo una exactitud de 1 y las muestras no asignadas 0.22; este error obtenido también aumenta con respecto al modelo anterior.

En la tabla 15 se tienen los valores obtenidos de especificidad, sensibilidad y precisión para el modelo tanto de calibración (*fitting*) y el de validación. Para cada caso y clase el valor obtenido fue de 1, lo que indica que este modelo también es óptimo para llevar a cabo la clasificación de las muestras por categoría.

Tabla 15. Valores de Especificidad, Sensibilidad y Precisión para modelo de calibración y validación en el intervalo de 850 a 1550 cm^{-1} .

Modelo	Clase	Especificidad	Sensibilidad	Precisión
Fitting	1	1	1	1
	2	1	1	1
Validación	1	1	1	1
	2	1	1	1

La matriz de confusión que se muestra en la tabla 16 indica que todas las muestras test han sido clasificadas y todas asignadas a la clase correcta (4 por cada clase).

Tabla 16. Matriz de confusión para el modelo construido en el intervalo de 1550 a 850 cm^{-1} .

	Clase 1	Clase 2	No asignados
Clase 1	4	0	0
Clase 2	0	4	0

La exactitud en la predicción de muestras fue de 1 en los dos modelos, con un error de 0, lo que nos indica que pese al aumento en el error de muestras no asignadas en el segundo modelo; éste realiza una asignación eficiente de las muestras test. Para esta técnica, ambos modelos presentan buen resultado por lo que ambos pueden ser utilizados para la determinación y clasificación de tequilas 100% agave y mixtos.

8.3.3 Vecino más cercano (KNN).

Para la elaboración del modelo por esta técnica se utilizó la matriz original con autoescalado como pretratamiento y sin tequilas anómalos en un intervalo de 1550 a 850 cm^{-1} . Para determinar este intervalo óptimo, primero se probó con la matriz original diferentes intervalos y tratamientos, siendo este intervalo el que dio los mejores resultados para los fines del modelo. Una vez elegida la base de datos, se obtuvo el valor de k el intervalo elegido como óptimo. El gráfico 29 muestra la tasa de error CV para cada valor de k . El valor de $k=3$ fue el que dio mejor resultado para la clasificación.

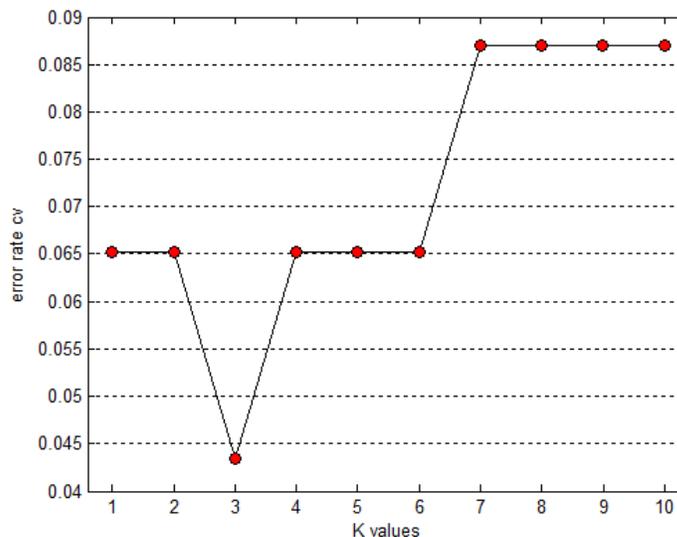


Figura 29. Tasa de error CV para cada valor de k .

La tabla 17 muestra los valores de los parámetros de especificidad, sensibilidad y precisión para los modelos de clasificación y predicción.

Tabla 17. Valores de Especificidad, Sensibilidad y Precisión para modelo de calibración y validación en el intervalo de 1550 a 850 cm^{-1} .

Modelo	Clase	Especificidad	Sensibilidad	Precisión
Fitting	1	0.91	1	0.94
	2	1	0.91	1
Validación	1	0.91	1	0.94
	2	1	0.91	1

Para el modelo de calibración y el conjunto de validación se obtuvo una exactitud de 0.96 con una tasa de error calculada de 0.04.

Tabla 18. Matriz de confusión para el modelo construido en el intervalo de 1550 a 850 cm^{-1} .

	Clase 1	Clase 2	No asignados
Clase 1	4	0	0
Clase 2	0	4	0

La predicción de las muestras test, asigna 4 muestras para cada clase y ninguna muestra sin asignar (tabla 18). Para esta técnica, este intervalo (850 a 1550 cm^{-1}) es el único que da como resultado la asignación correcta de las muestras test.

8.3.4 Analogía de clases por modelación suave e independiente (SIMCA).

La técnica de SIMCA, a diferencia de las otras técnicas, permite decidir el número de componentes a utilizar por clase. La tasa de error depende del número de componentes elegidos; no obstante, no siempre un mayor número de PCs se traduce en un mejor modelo, ya que una tasa de error menor, puede sobre ajustar el modelo.

Intervalo 1. 400 a 1550 cm^{-1} .

Para la matriz de calibración y validación se utilizó la misma matriz e intervalos previos. Los parámetros óptimos para cada intervalo se obtuvieron por medio de los gráficos de PC vs tasa de error y PC vs sensibilidad y especificidad. La figura 30 muestra el resultado de dichos gráficos para los datos de calibración en el intervalo de 400 a 1550 cm^{-1} .

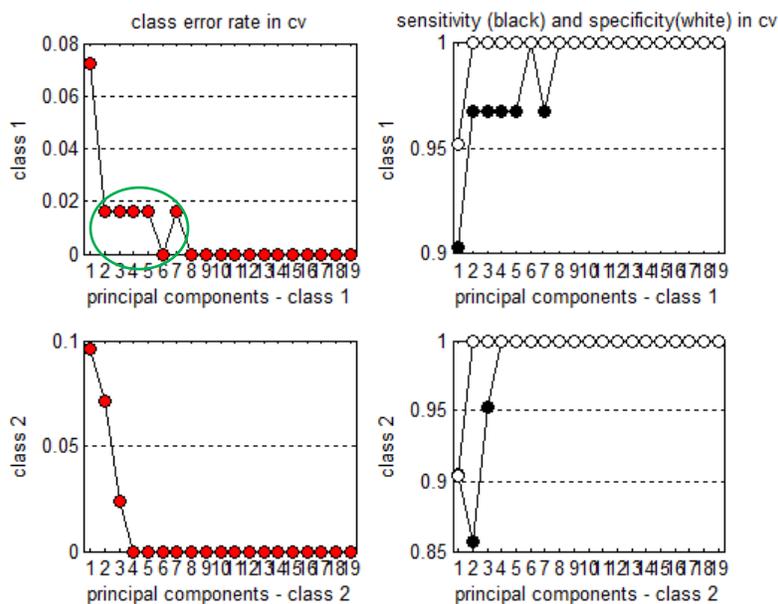


Figura 30. Gráficos PC vs Tasa de error y PC vs sensibilidad y especificidad para cada clase en el intervalo de 400 a 1550 cm^{-1} .

Los criterios para el cálculo de los gráficos mostrados en la figura 30 fueron los siguientes:

- Número de componentes principales: 10
- Escalado: Autoescalado
- Tipo de validación: Validación cruzada (persianas venecianas)
- Número de CV: 10
- Criterio de asignación: Modelado de clase

Con base en los resultados de los gráficos de PC vs tasa de error; fueron evaluados los PC de cada clase que presentan el menor valor de tasa de error, dando preferencia a los PC menores. El número de PCs que presentaron mejor resultado fueron para la clase 1, 6 PCs y para la clase 2, 4 PCs. Para la construcción del modelo de calibración se utilizaron los siguientes criterios:

- Escalado: Autoescalado
- Tipo de validación: Validación cruzada (persianas venecianas)
- Número de CV: 10
- Criterio de asignación: Modelado de clase
- Núm. De PCs Clase 1: 6
- Núm. PCs Clase 2: 4

La figura 31 muestra los valores para cada PC de la clase y su varianza explicada en el intervalo de 400 a 1550 cm^{-1} . PC1 para ambas clases es el componente que explica el mayor valor de porcentaje de la varianza explicada con 91.29 % para la clase 1 y 86.59 % para la clase 2. De manera general, el modelo explica el 99.99 % de la variación de las muestras, con una exactitud de 1 y tasa de error de 0.

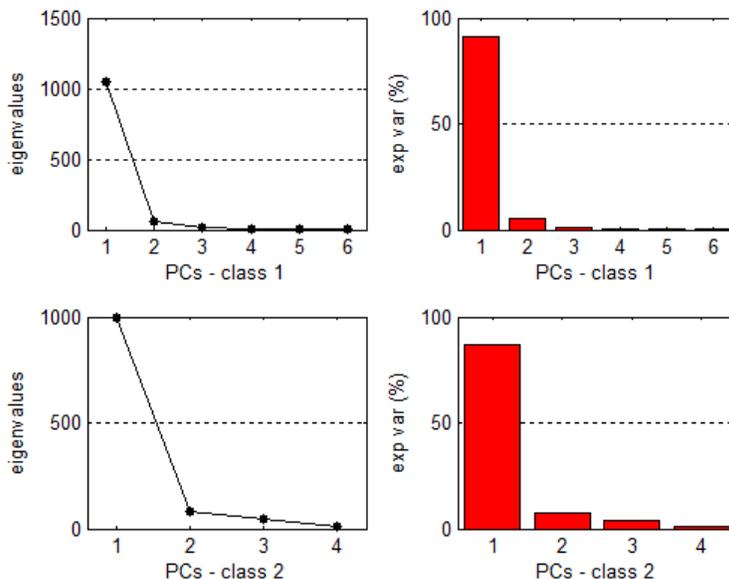


Figura 31. Gráficos de Valores por PC y porcentaje de varianza explicada para cada clase.

Una vez calculado el modelo se realizó la predicción de las muestras test obteniendo una exactitud de 1 y tasa de error de 0 para la validación; también el error para la asignación de muestras fue de 0 (consultar anexos). En la matriz de confusión de la tabla 19 observamos que el modelo es capaz de predecir las 8 muestras test de manera correcta asignándolas a su clase; también logra predecir con una exactitud de 1 y tasa de error de 0.

Tabla 19. Matriz de confusión para el modelo construido en el intervalo de 400 a 1550 cm^{-1} .

	Clase 1	Clase 2	No asignados
Clase 1	4	0	0
Clase 2	0	4	0

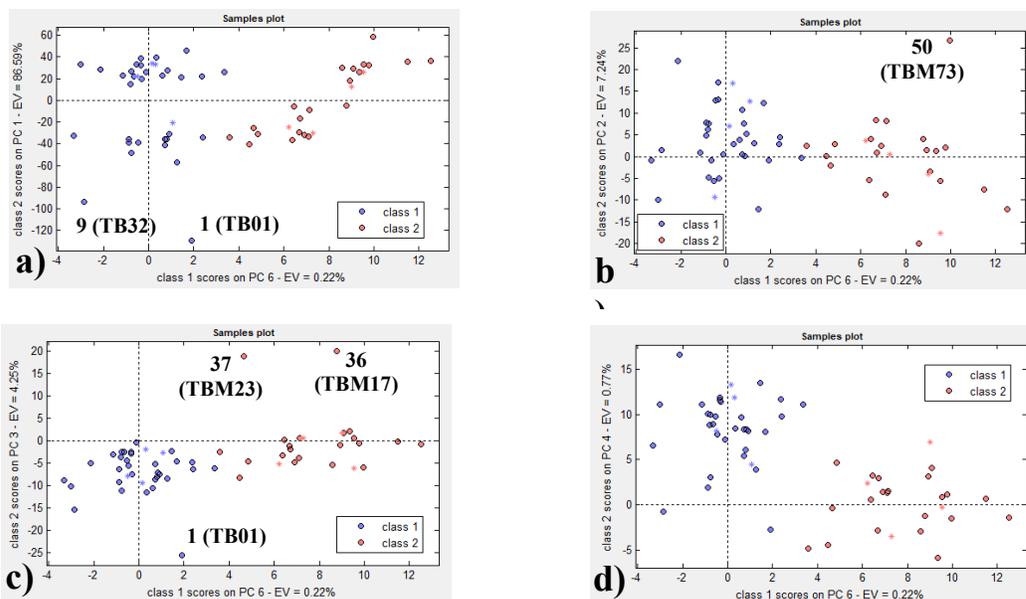


Figura 32. Gráficos de scores obtenidos en el intervalo de 1550 a 400 cm^{-1} . a) PC6 clase 1 vs PC1 clase 2; b) PC6 clase 1 vs PC2 clase 2; c) PC6 clase 1 vs PC3 clase 2; d) PC6 clase 1 vs PC4 clase 2.

Para los gráficos de scores de la figura 32 se tomó en cuenta cada PC de la clase 2 en el eje de las Y, mientras que en el eje de las X se mantuvo PC6 para la clase 1, ya que es a lo largo de este componente que la clase 1 se diferencia de la 2, pese a su bajo porcentaje de varianza.

La figura 32 a), al igual que en PCA, se observa a través de PC1 la formación de dos grupos divididos por el valor del eje Y igual a cero, lo que obedece a su distinto contenido alcohólico. En el resto de los gráficos desaparece este efecto al ya no intervenir PC1.

De manera general, en los gráficos de scores, tenemos muestras que se encuentran separadas de los grupos de su clase. Estos tequilas corresponden a las muestras TB01 y TB32, los cuales con base en lo explicado en sesión 7.2, se separan del resto por ser los que presentan mayor concentración de alcohol del grupo. En la figura 32 b) y c) se encuentran los tequilas TBM73 (tequila enriquecido con Frudex), TBM23 y TBM17, tequilas con menor concentración de alcohol del grupo.

En particular, observamos en el gráfico 32 b) y c) un acomodo horizontal sobre PC6 de la clase 1, con un mejor agrupamiento de ambas clases en este último gráfico, sobre todo para la clase 1 (tequilas 100 % agave). En el gráfico 32 d) (PC6 vs PC4) la separación es óptima y se infiere que ambos PCs aportan a la clasificación.

En todos los gráficos fue observable, de manera general, una mayor dispersión entre las muestras de los tequilas mixtos (clase 2) y esto en parte pudiera deberse a las diferentes fuentes de azúcar utilizadas para la fermentación, entre otras causas.

En general, fue posible observar una separación satisfactoria entre las clases, lo que indica que el modelo cumple para la diferenciación por categoría.

Análisis de Loadings (Intervalo 1).

Para determinar la influencia de los diferentes grupos funcionales del tequila, se calcularon los loadings para cada PC (figura 33). Al comparar éstos con los compuestos descritos en capítulos anteriores, se pudieron asignar algunos grupos a los loadings. Con fines prácticos, sólo se presentan los loadings de los PCs utilizados en los modelos.

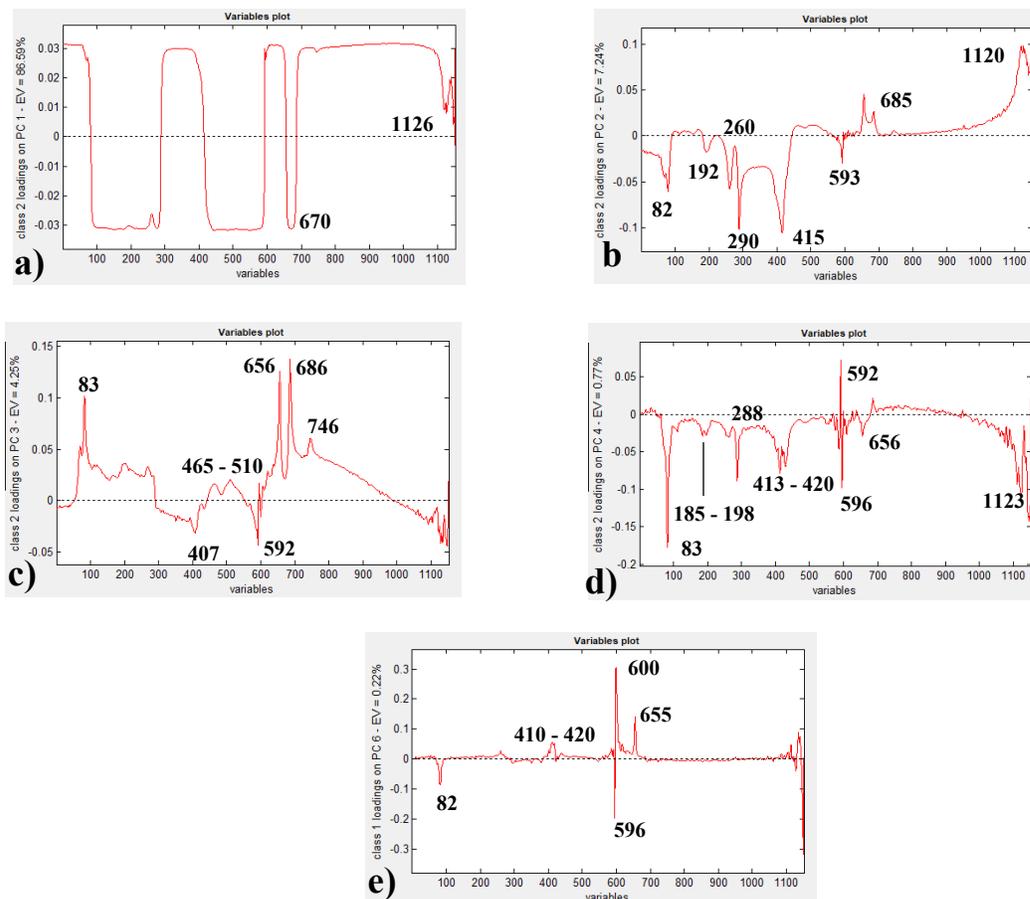


Figura 33. Gráficos de loadings obtenidos en el intervalo de 400 a 1550 cm^{-1} .
 a) Loading PC1 clase 2; b) Loading PC2 clase 2; c) Loading PC3 clase 2; d) Loading PC4 clase 2;
 e) Loading PC6 clase 1.

Para facilitar el análisis de los PCs, se realizó una tabla (tabla 20) con el número de onda a la que corresponde cada variable de los gráficos de la figura 33 y sus compuestos asignados a cada uno. En primera instancia para PC1 se observa bandas anchas, por lo que es difícil la asignación de algún compuesto en específico. Sin embargo, por la zona de las bandas, se puede inferir la influencia de señales complementarias de compuestos con grupos hidroxilo.

Tabla 20. Variables Matlab y su longitud de onda y compuestos asignados.

Longitud de Onda	Variable Matlab	Compuestos
1550	0	2-metil furil; 2,5-dimetil furano; propanol; 1-butanol, 3-metil acetato
1450	100	Furfural, benzaldehído; 3-metil butanol; ácido acético, etil éster; acetaldehído; 2-metil furil; 2,5-dimetil furano; isobutanol; 1-butanol, 3-metil acetato
1350	200	3-metil butanol; etil éster; acetaldehído; 2-metil furil; 2,5-dimetil furano; acetato 2-furanmetanol; isobutanol; 1-butanol, 3-metil acetato; oxalato
1250	300	Furfural, metanol; ácido acético; etil éster; 2,5-dimetil furano
1150	400	Furfural; fenil etanol; acetaldehído; 2-metil furil
1050	500	Furfural; 3-metil butanol; etil éster; 2,5-dimetil furano; propanol; isobutanol; 1-butanol, 3-metil acetato
950	600	Furfural; fenil etanol; ácido acético; etil éster; 2,5-dimetil furano; propanol; 2-fenil etil acetato
850	700	Etil éster; furfural;
750	800	Furfural; benzaldehído; etil éster; 2-metil furil; 2,5-dimetil furano; acetato 2-furanmetanol; 2-fenil etil acetato; oxalato
650	900	Benzaldehído; ácido acético; etil éster; 2-metil furil; propanol; 2-fenil etil acetato; oxalato
550	1000	2-metil furil; 2-fenil etil acetato
450	1100	Benzaldehído, metanol

Con base en la tabla 20, la aportación de PC2 a la varianza total viene dada principalmente por compuestos como el furfural y acetaldehído en la variable Matlab de 400 (1150 cm^{-1}); benzaldehído cerca de 100 (1450 cm^{-1}); etil éster y ácido acético cerca de 300 y 600; que son principalmente compuestos con grupo acetato como el 1-butanol 3-metil acetato y 2-fenil etil acetato que en técnicas cromatográficas se ha demostrado que son determinantes para la diferenciación por categoría (37).

En PC3 las señales aparecen en zonas similares que en PC2, pero a diferencia de los loadings en PC2, la forma de la señal cambia y son menos definidas. Estas señales van desde 192 a 415 con base en el gráfico 33 b). Estas señales también pueden ser asignadas a algunos compuestos furánicos, pero principalmente el grupo responsable es el etil éster.

El número de señales en los loadings de PC4 y PC6 es menor, pero mejor definidas que en los anteriores; éstas corresponden principalmente a compuestos con grupo éster, acetato y compuestos furánicos (400 a 500 correspondiente la zona de 1150 cm^{-1} a 1050 cm^{-1}), compuestos responsables de observar una diferencia entre ambas categorías.

Intervalo 2. 850 a 1550 cm^{-1} .

La matriz utilizada para el segundo modelo de calibración y validación fue la construida en el intervalo de 850 a 1550 cm^{-1} . Los parámetros óptimos para cada intervalo se obtuvieron calculando los gráficos de PC vs tasa de error y PC vs sensibilidad y especificidad. La figura 34 muestra el resultado de los gráficos para los datos de calibración.

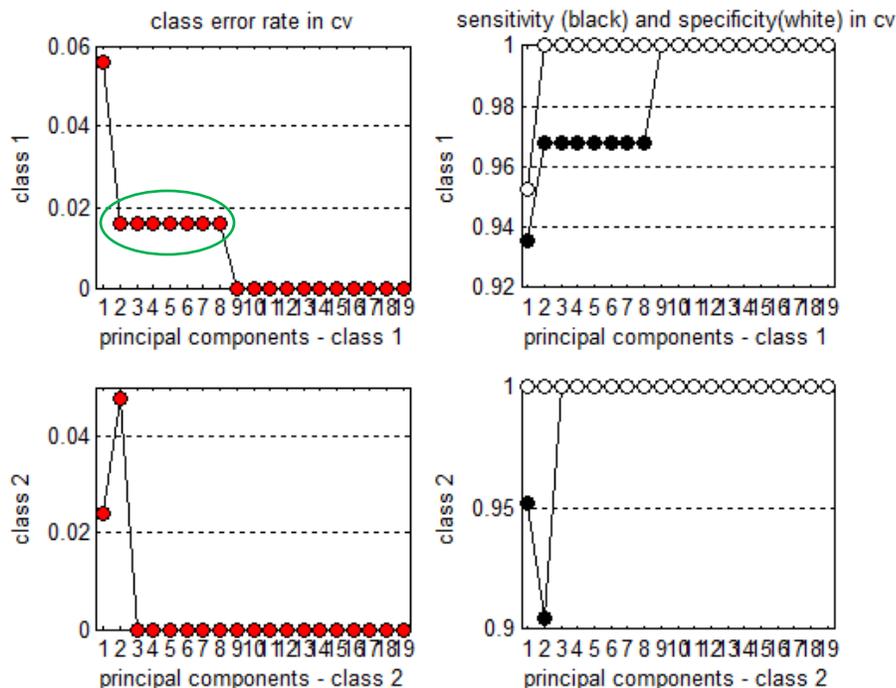


Figura 34. Gráficos PC vs Tasa de error y PC vs Sensibilidad y especificidad para cada clase de estudio.

Los criterios para el cálculo de los gráficos mostrados en la figura 34 fueron los siguientes:

- Número de componentes principales: 10
- Escalado: Autoescalado
- Tipo de validación: Validación cruzada (persianas venecianas)
- Número de CV: 10
- Criterio de asignación: Modelado de clase

Con base en los resultados de los gráficos de PC vs tasa de error, se eligió el número de PCs para cada clase; para el caso de la clase 1 se evaluó el modelo desde 2 hasta 4 PCs y para la clase 2 solo se tomó 3 componentes. El número de PCs elegido fue de 4 para la clase 1 y 3 para la clase 2. Para la construcción del modelo de calibración se utilizaron los siguientes criterios:

- Escalado: Autoescalado
- Tipo de validación: Validación cruzada (persianas venecianas)
- Número de CV: 10
- Criterio de asignación: Modelado de clase
- Núm. De PCs Clase 1: 4
- Núm. PCs Clase 2: 3

La varianza obtenida para el modelo fue del 99.9 9%; siendo así, este modelo también logra explicar el 100 % de la variación de las muestras. La figura 35 muestra los valores para cada PC de la clase y su varianza explicada. Al igual que el modelo anterior, PC1 para ambas clases es el componente que explica más porcentaje de la varianza explicada con 92.56 % para la clase 1 y 87.75 % para la clase 2. En el modelo de calibración se obtuvo una exactitud de 1 y una tasa de error de 0.

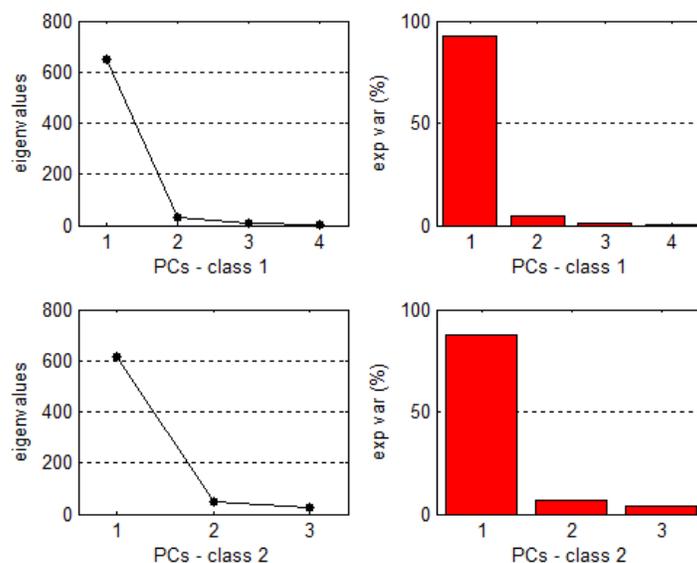


Figura 35. Gráficos de Valores por PC y porcentaje de varianza explicada para cada clase.

Una vez calculado el modelo, se realizó la predicción de las muestras test obteniendo una exactitud de 1 y tasa de error de 0 para la validación; también el error para la asignación de muestras fue de 0. En la matriz de confusión de la tabla 21 observamos que el modelo de nuevo es capaz de predecir las 8 muestras test de manera correcta asignándolas a su clase; el modelo logra predecir con una exactitud de 1 y tasa de error de 0.

Tabla 21. Matriz de confusión para el modelo construido en el intervalo de 400 a 1550 cm^{-1} .

	Clase 1	Clase 2	No asignados
Clase 1	4	0	0
Clase 2	0	4	0

Los gráficos de scores mostrados en la figura 36 fueron construidos utilizando los PCs de la clase 2 vs PC4 de la clase 1; siendo PC4 quien aporta su varianza para la diferenciación por categoría.

En la figura 36a (PC4 vs PC1), observamos que la diferenciación por categoría se da a lo largo del eje de las X; con la formación de dos grupos divididos por el eje $Y=0$. Arriba de este eje los grupos poseen un porcentaje menor de etanol que los inferiores; el mismo comportamiento fue observado en los modelos de PCA. Al igual que en los modelos anteriores, tenemos la separación de los tequilas TBM17, TBM23, TB01 y TB32 del resto de tequilas del grupo alto por su contenido alcohólico.

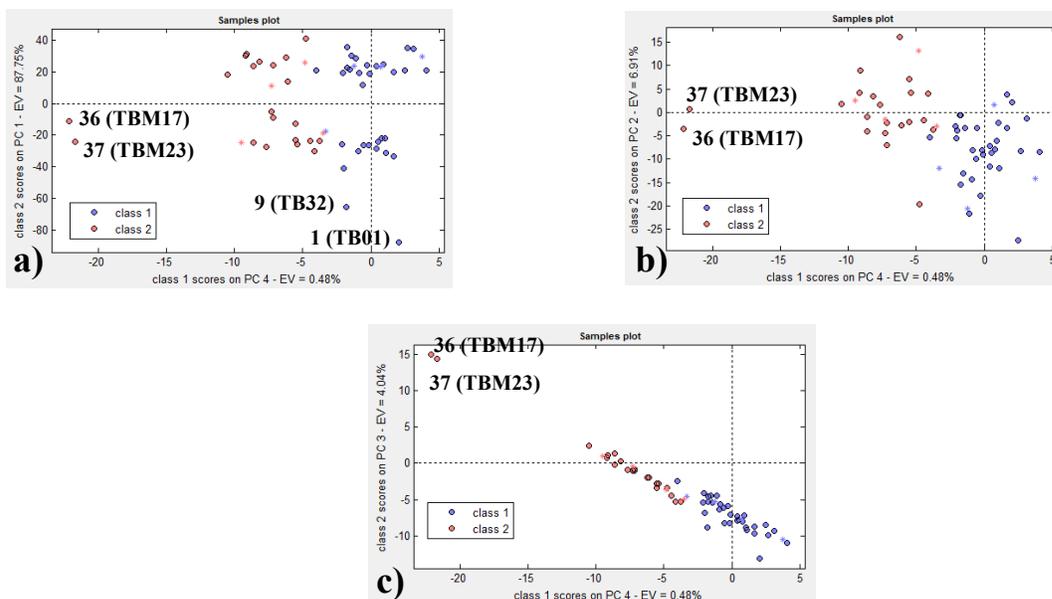


Figura 36. Gráficos de scores obtenidos en el intervalo de 1550 a 850 cm^{-1} .
 a) PC4 clase 1 vs PC1 clase 2; b) PC4 clase 1 vs PC2 clase 2; c) PC4 clase 1 vs PC3 clase 2.

Para la figura 36 b) la distribución de los grupos se da de manera cruzada en el eje PC4 vs PC2 con un mejor agrupamiento para los tequilas 100 % agave (clase1); los tequilas TBM17 y TBM23 aparecen al extremo de la izquierda separados del resto. En la figura 36 c) las clases se encuentran compactas en el extremo inferior derecho del diagrama, mientras que en la parte superior tenemos los tequilas con menor concentración de etanol del grupo.

Análisis de Loadings (Intervalo 2).

Para una mejor interpretación sobre la distribución de los scores de la figura 36, se obtuvieron los loadings (figura 37).

Para la explicación de los loadings se utilizó de nueva cuenta las señales y variables utilizada en el modelo anterior (tabla 22), sólo que esta la tabla se limita a las variables observadas en el eje X de 0 a 700 correspondientes al intervalo de 850 a 1550 cm^{-1} .

Para PC1 (figura 37 a) las señales por variable tienen una tendencia similar a la observada en el intervalo anterior y de igual forma se observan bandas anchas, por lo que no pueden asignar a grupo funcional determinado.

Para el resto de los PCs las señales son más definidas; para PC2 encontramos una señal intensa cerca de la variable 100, la cual está relacionada con compuestos con grupo acetato, éster y furano. En la variable 200 la principal contribución la podemos atribuir a grupos OH en alcoholes de cadena larga y el grupo oxalato, ya que se ha probado en técnicas cromatográficas diferencias en concentraciones de alcoholes como metanol e isobutanol entre tequilas mixtos y 100% agave lo cual podría atribuirse a la mayor disponibilidad de pectinas en los agaves para la fermentación (37).

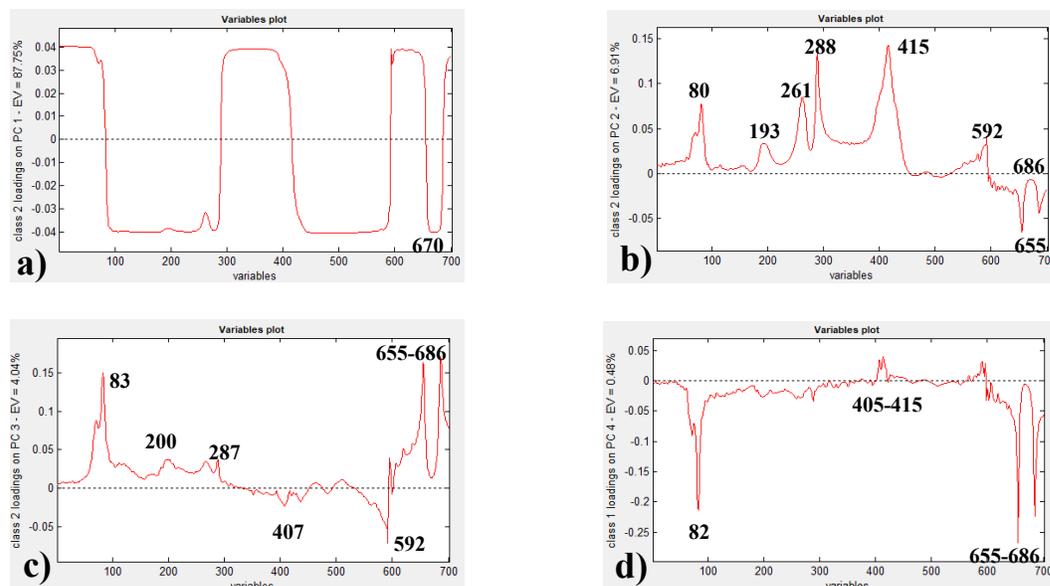


Figura 37. Gráficos de loadings obtenidos en el intervalo de 1550 a 850 cm^{-1} .

a) Loading PC1 clase 2; b) Loading PC2 clase 2; c) Loading PC3 clase 2; d) Loading PC4 clase 1.

Las variables en 300 y 400 (1250 y 1150 cm^{-1}), son las de mayor intensidad y son las que diferencian a este PC del resto; estas variables pueden ser asignadas a grupos con anillo aromático, como los derivados furánicos y fenoles. Las variables 600 y 700 (950 y 850 cm^{-1}) las podemos atribuir a grupos con grupo fenil, ácido acético, propanol y etil éster.

En los loadings de PC3 las señales principales se encuentran en los extremos (100 y 700). En la variable en 100 la señal tiene una tendencia similar a la observada en PC2 por lo que podemos inferir que se debe a los mismos compuestos (acetato, éster y furano). Por otro lado, en la variable de 700 las principales contribuciones son de acetaldehído, compuestos con alcoholes de cadena larga y acetatos, así como etil éster y el furfural.

Tabla 22. Variables Matlab y su longitud de onda y compuestos asignados

Longitud de Onda	Variable Matlab	Compuestos
1550	0	2-metil furil; 2,5-dimetil furano; propanol; 1-butanol, 3-metil acetato
1450	100	Furfural, benzaldehído; 3-metil butanol; ácido acético, etil éster; acetaldehído; 2-metil furil; 2,5-dimetil furano; isobutanol; 1-butanol, 3-metil acetato
1350	200	3-metil butanol; etil éster; acetaldehído; 2-metil furil; 2,5-dimetil furano; acetato 2-furanmetanol; isobutanol; 1-butanol, 3-metil acetato; oxalato
1250	300	Furfural, metanol; ácido acético; etil éster; 2,5-dimetil furano
1150	400	Furfural; fenil etanol; acetaldehído; 2-metil furil
1050	500	Furfural; 3-metil butanol; etil éster; 2,5-dimetil furano; propanol; isobutanol; 1-butanol, 3-metil acetato
950	600	Furfural; fenil etanol; ácido acético; etil éster; 2,5-dimetil furano; propanol; 2-fenil etil acetato
850	700	Etil éster; furfural;

Para los loadings de PC4 las señales son idénticas en cuanto a forma y tendencia a las observadas en PC3, pero de orden inverso por lo que deben estar relacionadas a los mismos compuestos. Poniendo esto en contexto con el gráfico de scores de la figura 36 c), al tener la contribución de los mismos compuestos en ambos ejes tenemos una distribución totalmente lineal y diagonal, lo que denota la contribución de ambos componentes.

Con estos resultados y con base en la varianza explicada para los modelos de ambos intervalos, se puede concluir que ambos son satisfactorios y permiten realizar la diferenciación de los tequilas por categoría.

Finalmente, la tabla 23 resume los parámetros de desempeño obtenidos para cada uno de los modelos de las técnicas supervisadas, los cuales fueron deducidos de la calibración y la validación en los intervalos de 450 a 1550 cm^{-1} y de 850 a 1550 cm^{-1} . En ambos casos, los resultados son satisfactorios. Los mejores parámetros de desempeño fueron los obtenidos en el primer intervalo. En calibración, los valores resultaron igual a la unidad, con todas las técnicas utilizadas, lo cual denota una excelente habilidad para la predicción de muestras pertenecientes a cada clase y rechazar otras. Los valores de exactitud y precisión fueron igual a 1. La validación resultó también satisfactoria, aunque sólo en el caso de KNN, la exactitud, la especificidad y la precisión son menores a 1, aunque aceptables. Las Funciones de potencia y SIMCA reportaron excelentes resultados tanto en calibración como en validación en ambos intervalos.

Tabla 23. Resumen de los parámetros de rendimiento para los diferentes modelos.

Método	Clase	450 – 1550 cm^{-1}							
		Calibración				Validación			
		Exac.	Espec.	Sens.	Prec.	Exac.	Espec.	Sens.	Prec.
FP	C1	1	1	1	1	1	1	1	1
	C2		1	1	1		1		1
KNN	C1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	C2								
SIMCA	C1	1	1	1	1	1	1	1	1
	C2		1	1	1		1	1	1
		850 – 1550 cm^{-1}							
		Calibración				Validación			
		Exac.	Espec.	Sens.	Prec.	Exac.	Espec.	Sens.	Prec.
FP	C1	1	1	1	1	1	1	1	1
	C2		1	1	1		1	1	1
KNN	C1	0.96	0.91	1	0.94	0.96	0.91	1	0.94
	C2		1	0.91	1		1	0.91	1
SIMCA	C1	1	1	1	1	1	1	1	1
	C2		1	1	1		1	1	1

N/A = No Aplica

Exac. = Exactitud

Espec. = Especificidad

Sens. = Sensibilidad

Prec. = Precisión

9. Conclusiones.

En el presente trabajo se construyeron modelos mediante técnicas no supervisada y supervisadas para la discriminación de los tequilas por categoría, 100 % agave y mixtos. Los 63 tequilas analizados pertenecieron sólo a la clase de los blancos. Los resultados obtenidos por técnicas no supervisadas, principalmente PCA y CA, permitieron establecer pautas sobre el comportamiento de las muestras de acuerdo con sus características inherentes, así como la identificación de muestras anómalas. Las muestras analizadas de tequila de ambas categorías fueron otorgadas por el CRT y por lo tanto, su origen, elaboración, lotes de producción y contenido alcohólico fue muy diverso. Las técnicas no supervisadas fueron PCA y CA, mientras que las técnicas supervisadas fueron Curvas de potencia, SVM, Funciones de potencia, KNN y SIMCA.

Por otra parte, se llevó a cabo la interpretación de un espectro típico de tequila a fin identificar los principales grupos funcionales, los cuales fueron comparados con los loadings de los modelos óptimos obtenidos por la técnica de PCA y así tener una idea de los principales componentes que contribuyen a diferenciar entre los tequilas de 100 % agave y mixtos. Estos grupos fueron algunos alcoholes de cadena larga tales como: isobutanol y butanol (relacionados con las señales observadas en PC3 para modelos como PCA, CA y CP); además también se encontró gran influencia de compuestos furánicos como el 2,5-dimetil furano y el furfural (relacionados con las señales de PC2) que en conjunto con el oxalato, el etil éster, el acetaldehído y benzaldehído (relacionados con las señales de PC2 y PC3) permitieron realizar la diferenciación en los modelos obtenidos por técnicas no supervisadas.

En particular, el gran margen de variación del porcentaje etanólico de las muestras representó un reto en la elaboración de los modelos ya que, en primera instancia, éstos evidenciaban dicha diferencia. No obstante, el ensayo de las condiciones (escalados, elección de intervalos de trabajo, eliminación de muestras anómalas, etc.) durante el desarrollo de los modelos, permitió alcanzar la optimización de los modelos, que en ningún caso fueron contradictorios sino confirmatorios. En general, los modelos propuestos en este trabajo proporcionaron resultados muy satisfactorios, dada la gran complejidad y diversidad de las matrices.

En los modelos obtenidos por técnicas supervisadas, la clasificación se logró por las tres técnicas utilizadas (KNN, FP y SIMCA). Por otro lado, la asignación de grupos a las señales observadas en técnicas como SIMCA dio un resultado similar a los obtenidos por técnicas no supervisadas. Los parámetros de desempeño obtenidos para la calibración y validación en cada una de las técnicas fueron adecuados para los dos intervalos probados, en todos los casos los valores de exactitud y precisión fueron igual a 1 en las técnicas de FP y SIMCA, y por arriba del 0.9 para KNN.

Si bien existen trabajos enfocados a la categorización de tequilas y bebidas alcohólicas por técnicas analíticas, el presente ensayo no solo aporta buenos resultados a este fin, también ofrece una metodología sencilla y rápida de análisis mediante estrategias quimiométricas que representan un parteaguas para futuras investigaciones enfocadas a la categorización de muestras más complejas (tequilas reposados y añejos).

Las condiciones establecidas para el pre-tratamiento de las muestras y los intervalos elegidos para cada técnica permitieron detectar componentes químicos específicos que diferencian las categorías, logrando así la clasificación por las técnicas propuestas. Como se ha mencionado, los

parámetros de trabajo y procesos establecidos para la clasificación, representan un aporte significativo y sientan bases para su uso en la industria del tequila, ofreciendo modelos sencillos que podrían ser usados en la detección de adulteraciones y tequilas 100 % agave falsos; no obstante, se requiere continuar trabajando con más detalle en la caracterización de compuestos químicos presentes en los tequilas, fin que se sale del alcance de este trabajo, pero que permite develar la importancia de un buen tratamiento de muestras y una selección adecuada del intervalo de trabajo.

10. Bibliografía

10.1 Referencias

1. NOM-006-SCFI-2012. (2012). *Bebidas alcohólicas-Tequila-Especificaciones*. Diario Oficial de la Federación. México.
2. García-Montiel, R. (2015). *Espectrometría Infrarroja Media (MIR) y Reconocimiento de Pautas para la Diferenciación de Tequilas*. Facultad de Estudios Superiores Cuautitlán. UNAM. México.
3. Tomás i Morer, X. y Andrade Garda, J.M. (1999). *Aplicación de las curvas de potencial simplificadas a problemas de clasificación*. Química Analítica, 18 (3), 225-231.
4. Mateos García, D. (2013). *Ponderación local evolutiva de la regla KNN*. Tesis de Doctorado. Universidad de Sevilla, 7-14.
5. Pérez-Caballero, G., Andrade, J.M., Olmos, P., Molina, Y., Jiménez, I., Durán, J.J., Fernández-Lozano, C. Miguel-Cruz, F. (2017). *Authentication of tequilas using pattern recognition and supervised classification*. Trends in Analytical Chemistry, 94, 117-129.
6. Massart, D.L., Vandeginste, B.G.M., Buydens, L.M.C., de Jong, S., Lewi, P.J. y Smeyers-Verbeke, J. (1997). *Handbook of Chemometrics and Qualimetrics: Part A and B*. Amsterdam, Elsevier.
7. Arvanitoyannis, I., Katsota, M., Psarra, E., Soufleros, E.H. y Kallithraka, S. (1999). *Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics)*. Trends in Food Science and Technology, 10, 321.
8. Zayas, E.P., Lorenzo, M., Hernández, D. y Blanco, I. (2014). *Clasificación multivariante de rones añejos cubanos*. Revista Cubana de Ingeniería, 2, 62-67.
9. U. Hener, A. Mosandl, A. Hilkert, J. Bahrs-Windsberger, M. GroBmann, W.R. Sponholz. (1998). *Detection and Partial Characterisation of New Anthocyanin-Derived Pigments in Wine*. Vitic En01 Sci 2, 1-11.
10. G.E. Krammer, I. Gatfield, M. Güntert, J.M. Hilmer, C.O. Schmidt, H. Sommer, P. Werkhoff, J. Kaulen, Schieberle P, Engel KH. (2001). *Frontiers of flavour science*. DFA, Garching, 111-116.
11. Y. Henry, C.B. Douthitt, D. Juchelka. (2003). *Authentication of tequila by gas chromatography and stable isotope ratio analyses*, Poster at the sixth International Symposium on Food Authenticity and Safety.
12. B.O. Aguilar-Cisneros, M.G. López, E. Richling, F. Heckel, P. Schreier. J. (2002). *Tequila Authenticity Assessment by Headspace SPME-HRGC-IRMS Analysis of ¹³C/¹²C and ¹⁸O/¹⁶O Ratios of Ethanol*, Agric Food Chem 50, 7520-7523
13. G. Martin, J.S. Akoka, M.L. Martin. (2006). *SNIF-NMR—Part 1: Principles*. G.A. Webb (Ed.), Modern magnetic resonance, part 3, Springer, Dordrecht, Netherlands 1629
14. C. Bauer-Christoph, N. Christoph, B.O. Aguilar-Cisneros.; M.G. López, E. Richling, A. Rossmann, P. Schreier. (2003). *Authentication of tequila by gas chromatography and stable isotope ratio analyses*. Eur. Food Res. Technol. 217, 438–443.
15. Casa Sauza. (2018). Página Inicial. México. <http://www.casasauza.com>.
16. Consejo Regulador del Tequila. (2018). Página Inicial. México. <https://www.crt.org.mx>.

17. EMBAMEX. (2018). Tequila, la bebida Mexicana Favorita del Mundo. México. <https://embamex.sre.gob.mx/paisesbajos/index.php/comunicados-boletines/170-tequila>.
18. Mancilla N.A y López M.G. (2002). *Generation of Maillard Compounds from Inulin during the Thermal Processing of Agave tequilana Weber Var. Azul*. Agric. Food Chem. 50, 806-812.
19. Díaz-Montaña D, Marie-Line D, Estarrón-Espinoza M y Strehaiano P. (2008). *Fermentative capability and aroma compound production by yeast strains isolated from Agave tequilana Weber juice*. Enzyme and Microbial Technology. 42, 608-616.
20. U. Contreras-Loera, O. Barbosa-García, G. Ramos Ortiz, J.L. Pichardo-Molina1, M.A. Meneses-Naval, J.L. Maldonado. (2009). *Identificación y discriminación de Tequilas reposados in situ para la protección de marca*. Revista Electrónica Nova Scientia. Nº 2 Vol. 1 (1). Pp 22-32.
21. López Nieves M.R. (2015). *Identificación y Discriminación de Tequilas 100% agave mediante la Aplicación de la Espectrofotometría Ultravioleta-Visible y de Técnicas Quimiométricas Multivariantes*. Facultad de Estudios Superiores Cuautitlán. UNAM. México.
22. Ramis Ramos G., Álvarez-Coque C, (2003), *Quiometría*, Madrid España, Editorial Síntesis.
23. Mongay Fernández, (2005), *Quiometría*, Universidad de Valencia España, Editorial PUV.
24. Swartz E. M., Krull S. I., (1997), *Analytical Method Development and Validation*, New York USA, Editorial Marcel Dekker.
25. Otto M., (1999), *Chemometrics: Statistic and Computer Application in Analytical Chemistry*, Alemania, and Editorial Wiley-VCH.
26. Skoog A. D., (2007), *Principles of Instrumental Analysis*, Sexta edición, USA, Editorial Books/Cole, Cengage Learning.
27. CCIQS UNAM, (2018). *Espectroscopía de Infrarrojo/ATR*, México. http://www.cciqs.unam.mx/index.php?option=com_content&view=article&id=113&Itemid=86.
28. Pérez-Caballero, G, (2015). *Espectroscopia Infrarrojo por transformadas de Fourier (FTIR) [PowerPoint]*. FESC UNAM, México.
29. Miller, J., (2002), *Estadística y quimiometría para química analítica*, Madrid, España. Editorial Prentice Hall.
30. Beebe R. Kenneth, Pell J. Randy, Mary Beth Seasholtz, (1998), *Chemometrics a practical guide*, USA, Wiley-Interscience.
31. Rouessac Francis, (2003), *Análisis Químico Métodos y Técnicas Instrumentales Modernas*, España, McGraw-Hill.
32. Blanco Marcel, (2007), *Temas Avanzados de Quiometría*, Palma.
33. Gonzales C., (2005), *Control Estadístico Multivariable mediante Análisis de Componentes Principales (PCA), aplicado al área de destilación Planta I Methanex Chile Limited*, Universidad de Magallanes.

34. Gutiérrez P. Anna. (2010), *La espectroscopia NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabados*, Departamento de Química, Universidad Autónoma de Barcelona, España.
35. Gutiérrez Ricardo. (2012), *Aplicación de Métodos Quimiométricos para la Caracterización y Control de Calidad de Plantas Medicinales*. Departamento de Química, Universidad Autónoma de Barcelona, España.
36. Macho P. Santiago. (2002), *Metodologías Analíticas Basadas en Espectroscopia de Infrarrojo y Calibración Multivariante. Aplicación a la Industria Petroquímica*. Área de Química Analítica, Universidad de Rovira.
37. Lachenmeier W. Dirk, Richling Elke, Lopéz G. Mercedes. (2005), *Multivariate Analysis of FTIR and Ion Chromatographic Data for the Quality Control of Tequila*. Journal of Agricultural and Food Chemistry. 53, 2151-2157.
38. Salinas, M. L.; Ogura, T.; Soffchi, L. (2001). *Irritant contact dermatitis caused by needle-like calcium oxalate crystals, raphides, in agave tequilana among workers in tequila distilleries and agave plantations*. Contact Derm., 44, 94-96.
39. Velazquez de la Cruz G. (2000), *Cuantificación por IR del agua directamente unida a una matriz polimérica*. Sociedad Mexicana de Ciencia de Superficies y de Vacío. 11, 24-31.

10.2 Índice de imágenes

[Imagen 1]

(2018). Cómo se hace el tequila. Recuperado de <http://www.casasauza.com/fabricacion-tequila>.

[Imagen 2]

Jasso C. (2018). Fotogalería “La cosecha de la plantación de tequila, en imágenes”. Recuperado de https://www.clarin.com/mundo/cosecha-plantacion-tequila-imagenes_0_SJk6tuDTM.html.

[Imagen 3]

(2018). Elaboración y Normativa. Recuperado de <https://www.crt.org.mx>.

[Imagen 4]

Arreola I. (2018). Denominación de Origen “Tequila”. Recuperado de <http://www.casasauza.com/todo-sobre-tequila/denominacion-de-origen-tequila>.

[Imagen 5]

(2018). Denominación de Origen del Tequila. Recuperado de https://www.crt.org.mx/index.php?option=com_wrapper&view=wrapper&Itemid=408.

[Imagen 6]

Pérez C. (2015). Espectroscopia Infrarrojo por Trasformadas de Fourier (FTIR).

[Imagen 7]

Departamento de Fisicoquímica UNAM. Fuerzas Intermoleculares. Recuperado de http://depa.fquim.unam.mx/amyd/archivero/11.Fuerzasintermoleculares_23730.pdf

[Imagen 8]

Fernández G. Fundamentos de Espectroscopía Infrarroja. Recuperado de <http://www.quimicaorganica.org/determinacion-estructural/160-espectroscopia-de-infrarrojo.html>.

[Imagen 9 y 10]

Skoog A. (2007). Espectrometría de absorción en el infrarrojo. Recuperado de Principes of Instrumental Analysis, Sexta edición.

[Imagen 11]

Universidad de Alcalá. Espectroscopía Vibracional. Recuperado de http://www3.uah.es/edejesus/resumenes/DECI/tema_2.pdf.

[Imagen 12]

Skoog A. (2007). Espectrometría de absorción en el infrarrojo. Recuperado de Principes of Instrumental Analysis, Sexta edición.

[Imagen 13 y 14]

Macho P. Santiago. (2002). Espectroscopia de Infrarrojo. Recuperado de Metodologías Analíticas Basadas en Espectroscopia de Infrarrojo y Calibración Multivariante. Aplicación a la Industria Petroquímica

[Imagen 15, 16 y 17]

Pérez C. (2015). Espectroscopia Infrarrojo por Transformadas de Fourier (FTIR).

[Imagen 18]

Piketech. (2018). ATR Theory & Application. Recuperado de www.piketech.com.

[Imagen 19, 20 y 21]

Macho P. Santiago. (2002). Espectroscopia de Infrarrojo. Recuperado de Metodologías Analíticas Basadas en Espectroscopia de Infrarrojo y Calibración Multivariante. Aplicación a la Industria Petroquímica

[Imagen 22]

Gonzales C. (2005). Control Estadístico Multivariable mediante Análisis de Componentes Principales (PCA), aplicado al área de destilación Planta I Methanex Chile Limited. Recuperado de http://www.umag.cl/biblioteca/tesis/gonzalez_carrasco_2005.pdf.

[Imagen 23]

Fernandez S. (2011). Análisis Conglomerados. Recuperado de <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>.

[Imagen 24]

Fraga Monfort C. (2014). Validación y transferencia de métodos analíticos mediante espectroscopia NIR. Recuperado de https://ddd.uab.cat/pub/tfg/2014/136795/TFG_CristinaMonfortFraga.pdf.

[Imagen 25]

Ramis Ramos G., Álvarez-Coque C, (2003). Análisis Clasificadorio. Recuperado de *Quiometría*.

[Imagen 26 y 27]

Matthias Otto, (1999). Chemometrics.

11. Anexos

11.1 Abreviaturas

CVPDB de Deuterio	Cromatografía de gases y espectrometría de masas por relación isotópica
FT-MIR	Infrarrojo Medio con Transformadas de Fourier
CRT	Consejo Regulador del Tequila
CATR	Corrección de ATR (Reflexión Total Atenuada)
CILB	Corrección Interactiva de la Línea Base
PCA	Análisis de Componentes Principales
CA	Análisis Cluster
SIMCA	Análisis de Clase por Modelado Suave e Independiente
PC	Curvas de Potencia
PF	Funciones de Potencia
K-NN	Vecino más Cercano
ATR	Reflexión Total Atenuada
EMA	Entidad Mexicana de Acreditación
DGN	Dirección General de Normas
COFEPRIS	Comisión Federal para la Protección contra Riesgos Sanitarios
LPI	Ley de Protección Industrial
DR	Reflexión Difusa (Diffuse Reflection)
USP	United States Pharmacopeia Exac = Exactitude
Esp	Especificidad
Sens	Sensibilidad
Prec	Precisión

11.2 Definiciones

Abocado.- Procedimiento para suavizar el sabor del Tequila, mediante la adición de uno o más de los siguientes ingredientes: Color Caramelo, Extracto de roble o encino, glicerina o jarabe a base de azúcar.

Cross-validación.- Utilización de los datos del conjunto de calibración para validar el modelo. En la cross-validación, el conjunto de calibración inicial se divide en diferentes grupos de cancelación. Cada vez que se calcula el modelo se deja un grupo de cancelación fuera y se calcula la ecuación de la regresión con los demás datos. Una vez realizada la calibración se testa el modelo con los datos del grupo de cancelación excluido.

Especificidad.- Es la habilidad del modelo para reconocer al analito de interés, muestra o grupo de muestras de manera precisa y específica en presencia de otros componentes en una matriz.

Exactitud.- Aproximación entre el resultado medio, obtenido de aplicar reiteradamente el procedimiento experimental un gran número de veces. En otras palabras, un resultado exacto es lo que en estadística se denomina in-sesgado.

Frecuencia.- Número de vibraciones completas por unidad de tiempo. Se representa con el símbolo “ ν ” y sus unidades son los Hertz (Hz).

Loading.- Variables (J) por cada muestra analizada (I). En un análisis de componentes principales en el cual se proporciona una aproximación de la matriz R, los loadings son los que retienen la relación existente entre las variables.

Longitud de onda.- Distancia medida en dirección de la propagación de la onda entre las dos partículas más próximas, que se encuentran en el mismo instante en una fase o estado de vibración. Se simboliza con: λ , las unidades utilizadas son micrómetros ($\mu\text{m}=10^{-6}\text{m}$), nanómetros ($\text{nm}=10^{-9}\text{m}$) y angstroms ($\text{\AA}=10^{-10}\text{m}$).

Momento Dipolar.- Interacciones que se dan entre las moléculas covalentes y que permiten que éstas se mantengan unidas; representa la distribución de la densidad electrónica en un enlace.

Mosto.- Líquido azucarado, proveniente de la extracción del agave hidrolizado y adicionado de otros azúcares conforme a la NOM-006-SCFI-2012, listo para el proceso de fermentación.

Número de onda.- Inversa de la longitud de onda. Las unidades más utilizadas son: ondas/cm = cm^{-1} (Káiser).

Piña.- Se denomina “piña” al tallo y la parte basal de las pencas de agave las cuales se obtienen después de un proceso de jima en el cual se separa la parte no aprovechable de la planta. En esta parte del agave es donde se encuentran almacenados los carbohidratos utilizados para la elaboración del tequila.

Precisión.- Medida del grado de repetibilidad de un método analítico en condiciones normales.

Pretratamiento.- Manipulaciones matemáticas aplicadas a una muestra o matriz de datos antes de cualquier otro tipo de análisis; consiste en anular y/o reducir fuentes de variabilidad en la señal, ya sea de carácter aleatorio (ruido) o de carácter sistemático (variaciones en la línea base, etc.), que no están relacionadas con el analito o la propiedad de interés.

Repetibilidad.- Es el valor por debajo del cual se está situado, con una probabilidad especificada, el valor absoluto de la diferencia entre dos resultados individuales obtenidos en las

mismas condiciones (mismo operador, mismo equipo, mismo laboratorio y mismo intervalo de tiempo).

Reproducibilidad.- Es el valor por debajo del cual se está situado, con una probabilidad especificada, el valor absoluto de la diferencia entre dos resultados individuales obtenidos en condiciones diferentes (diferente operador, diferente equipo, diferente laboratorio y diferente intervalo de tiempo).

Robustez.- Con base en la USP, la robustez es el grado de reproducibilidad de los resultados obtenidos bajo ciertas condiciones. Dentro de las condiciones de trabajo se considera el laboratorio, analista, periodo en el que se desarrolla el experimento (fechas), reactivos y equipos. Es expresado como desviación estándar relativa (%RSD).

Varianza explicada.- Varianza del vector, si el vector es un componente principal se denomina "Autovalor". Varianza de las proyecciones de los puntos sobre el vector. Se calcula como la suma de cuadrados de las distancias de los puntos al centroide en la dirección del vector dividido entre "n".

Varianza explicada acumulada.- Porcentaje de información contenido en el modelo con respecto a la información de la matriz de datos.

Varianza residual.- Suma de cuadrados de las distancias de los puntos al modelo dividido entre "n".

Varianza total (S_T^2).- Suma de los cuadrados de las distancias de los puntos a su centroide dividida por el número de puntos "n".

11.3 Reactivos

Tabla 24. Reactivos utilizados en la experimentación.

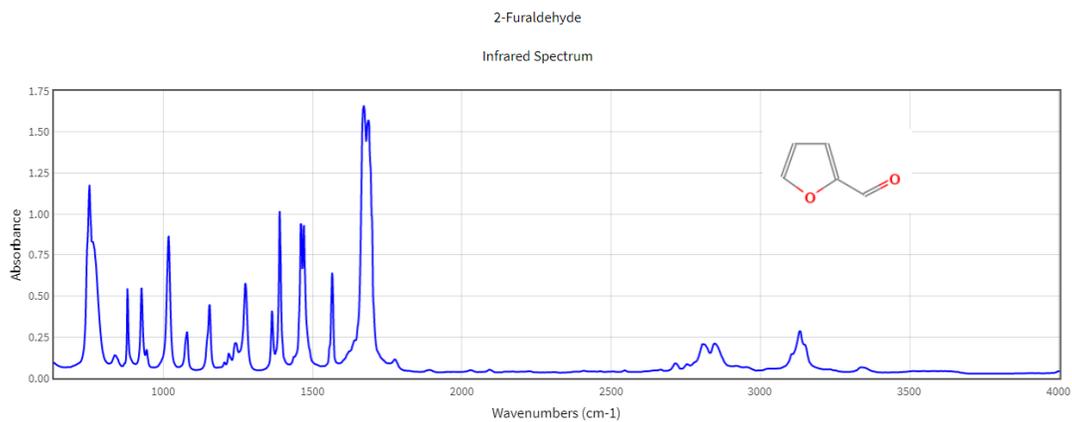
Nombre	Fórmula	Propiedades y características
Etanol absoluto	C_2H_6O	Líquido incoloro Peso molecular: 46.07 g/mol Pureza: 100%
Etanol anhidro	C_2H_6O	Líquido incoloro Peso molecular: 46.07 g/mol Pureza: 99.9%

11.4 Espectros MIR de componentes del tequila.

2-Furfuraldehído

Condiciones: LIQUID (NEAT)

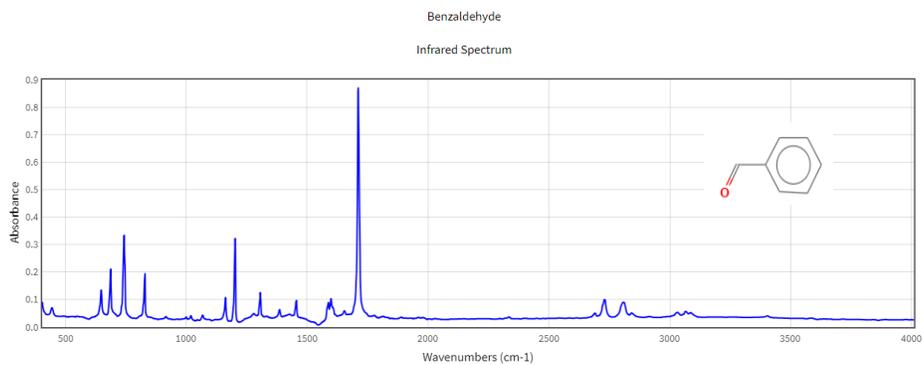
Instrumento: PERKIN-ELMER 521 (GRATING)



Benzaldehído

Condiciones: SOLUTION (1% IN CCl₄ FOR 3800-1330, 1% IN CS₂ FOR 1330-450 CM⁻¹) VS SOLVENT

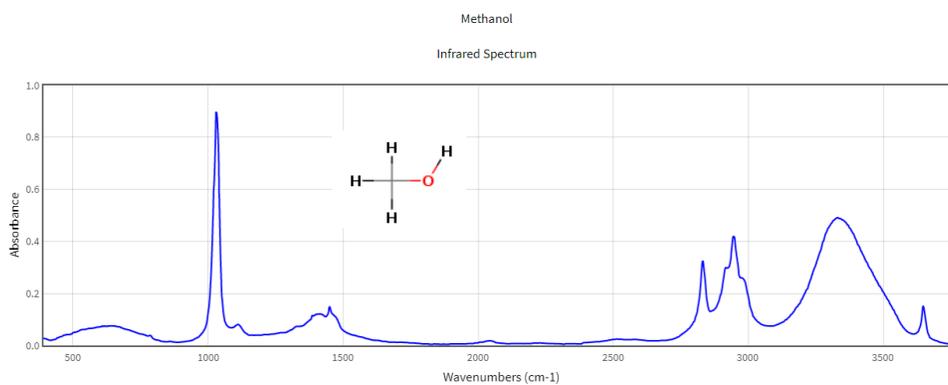
Instrumento: PERKIN-ELMER 180



Metanol

Condiciones: SOLUTION (2% IN CCl₄ FOR 3800-1330, 2% IN CS₂ FOR 1330-400 CM⁻¹) VS SOLVENT

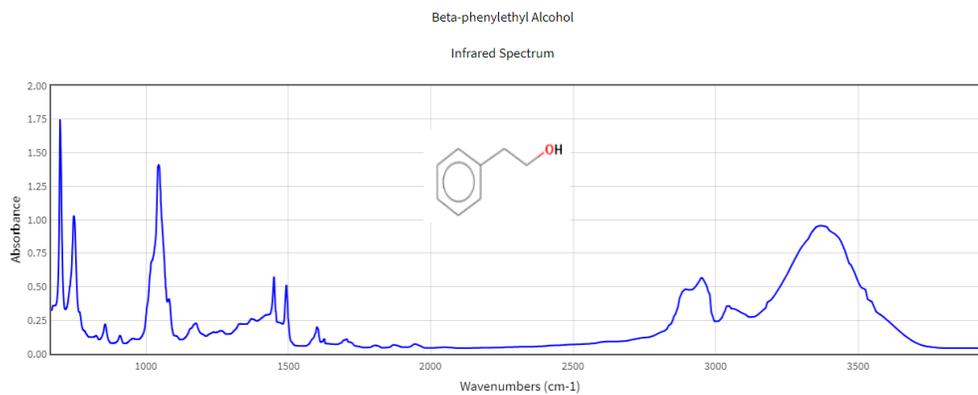
Instrumento: DOW KBr FOREPRISM-GRATING



Fenil etanol

Condiciones: LIQUID 96% PURE

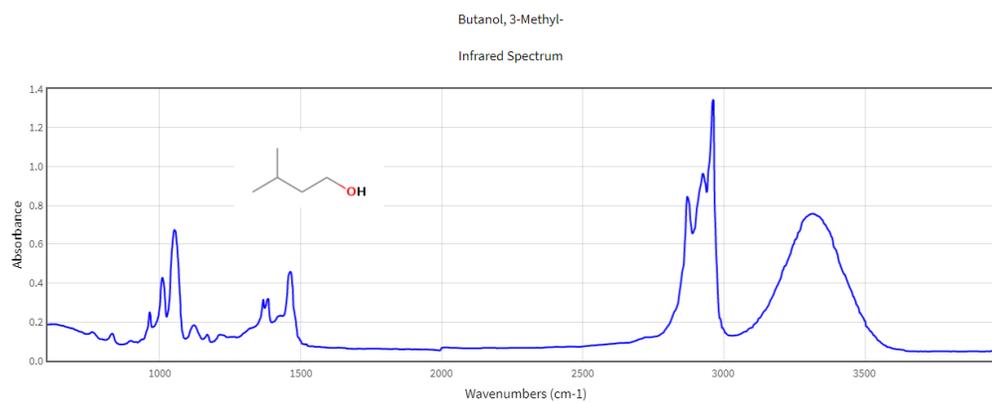
Instrumento: Not specified, most likely a prism, grating, or hybrid spectrometer.



3-Metil Butanol

Condiciones: LIQUID (NEAT) PURITY - 99%

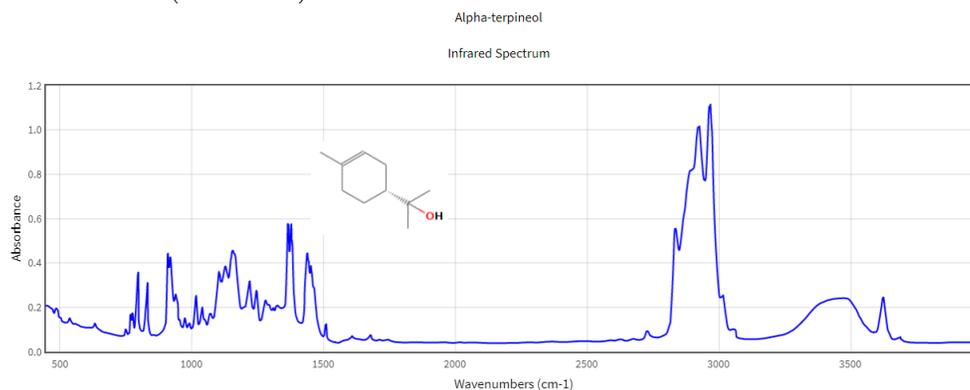
Instrumento: PERKIN-ELMER 421 (GRATING)



Alfa Terpineol

Condiciones: SOLUTION (10% CCl₄ FOR 4000-1330, 10% CS₂ FOR 1330-450 CM-1)

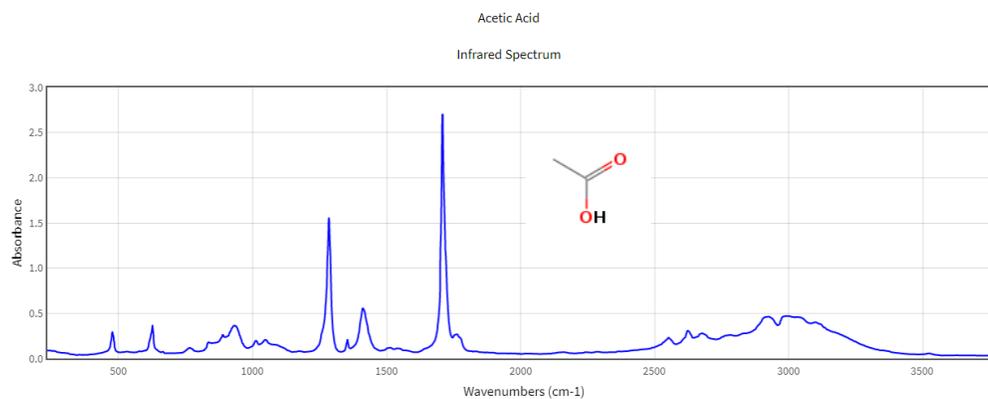
Instrumento: CARY 90 (GRATING)



Ácido Acético

Condiciones: SOLUTION (10% IN CCl₄ FOR 3800-1300, 10% IN CS₂ FOR 1300-650, 10% IN CCl₄ FOR 650-250 CM-1) VERSUS SOLVENT

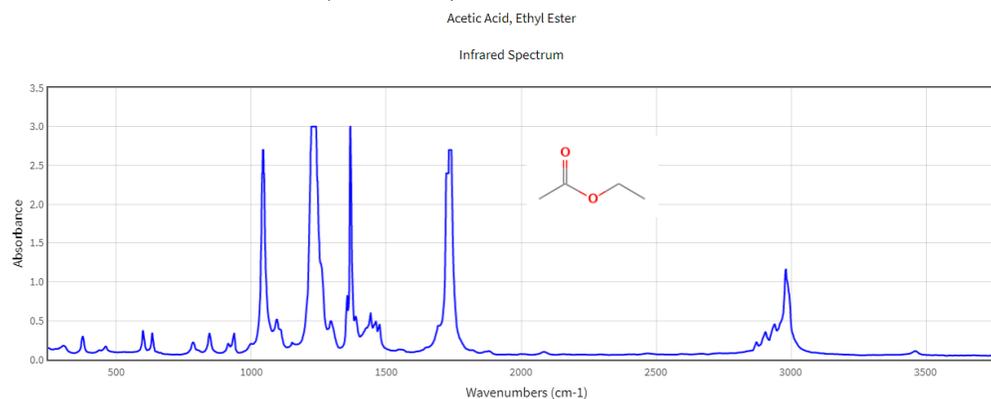
Instrumento: PERKIN-ELMER 521 (GRATING)



Acetato de etilo

Condiciones: SOLUTION (10% IN CCl₄ FOR 3800-1300, 10% IN CS₂ FOR 1300-650, 10% IN CCl₄ FOR 650-250 CM-1) VERSUS SOLVENT

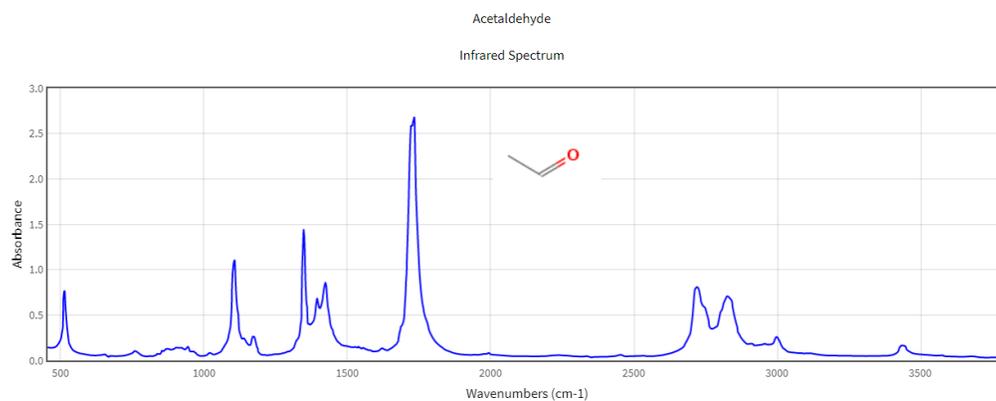
Instrumento: PERKIN-ELMER 521 (GRATING)



Acetaldehído

Condiciones: SOLUTION (10% CCl₄ FOR 2.6-7.5, 10% CS₂ FOR 7.5-22)

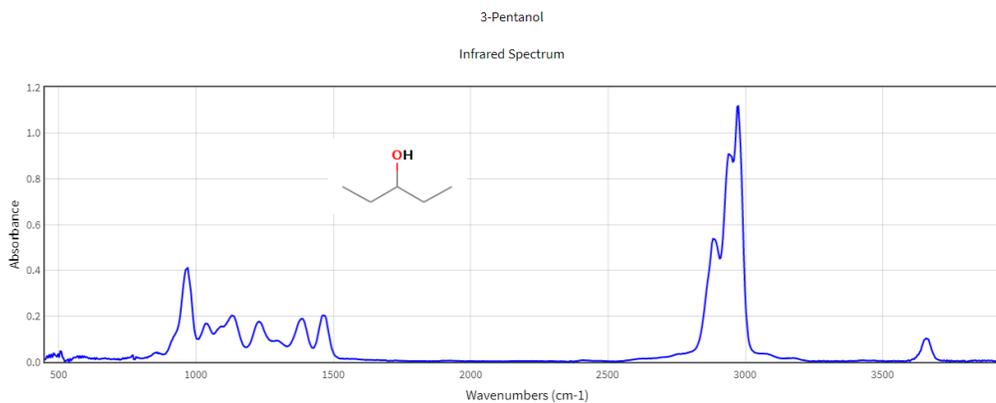
Instrumento: DOW KBr FOREPRISM-GRATING SPECTRAL CONTAMINATION DUE TO CCl₄ AROUND 1550 CM-1



3-Pentanol

Condiciones: Sin información

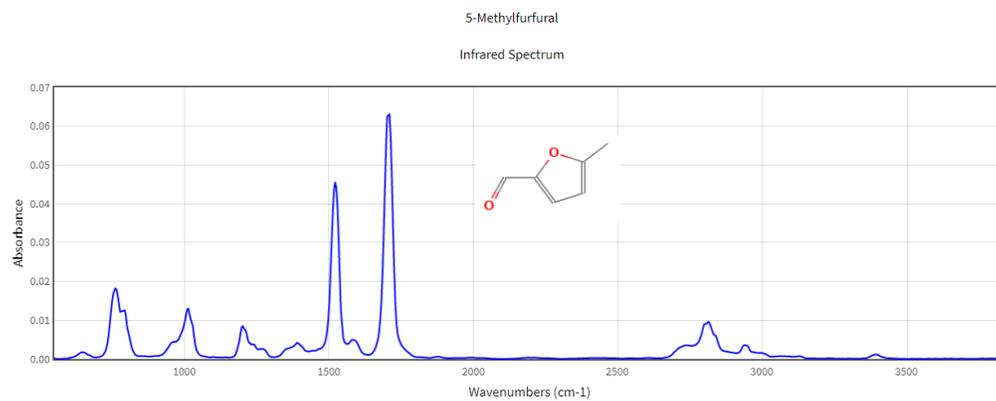
Instrumento: Sin información



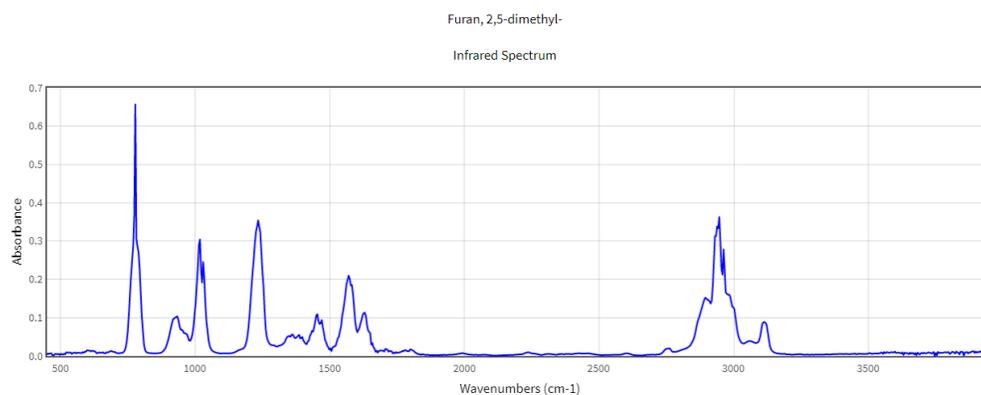
5-Metil furfural

Condiciones: Sin referencias

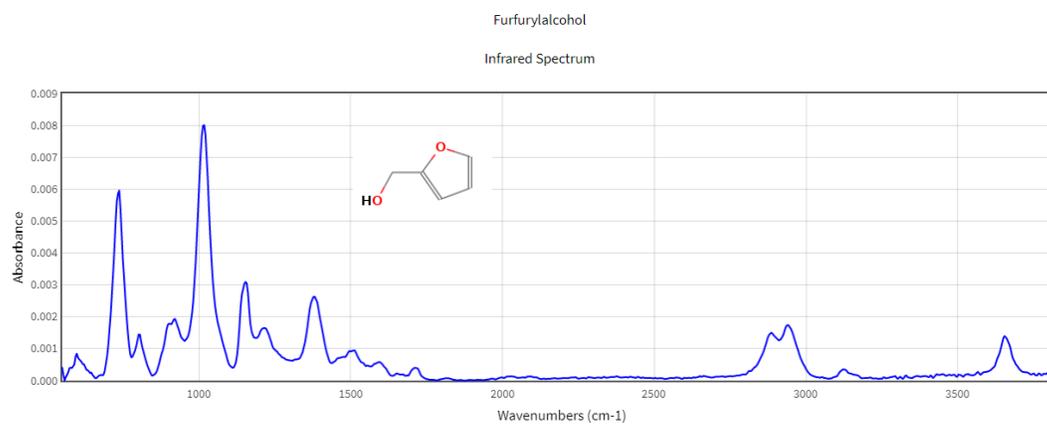
Instrumento: Sin referencias



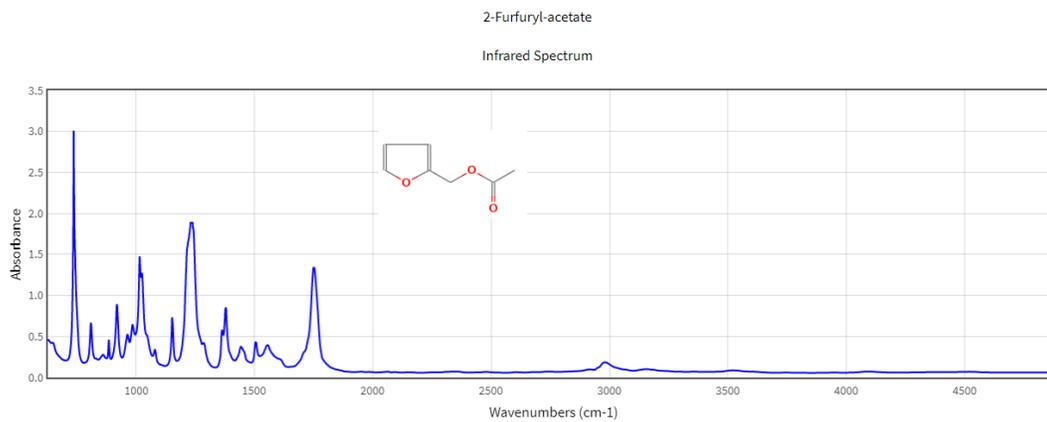
2,5-Dimetil Furano

Condiciones: Sin referencias**Instrumento:** Sin referencias

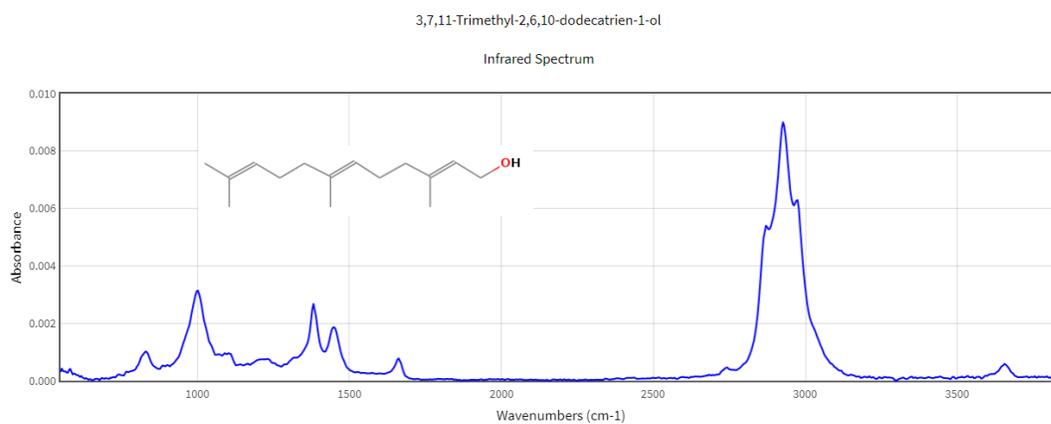
2-Furan metanol

Condiciones: Sin referencias**Instrumento:** Sin referencias

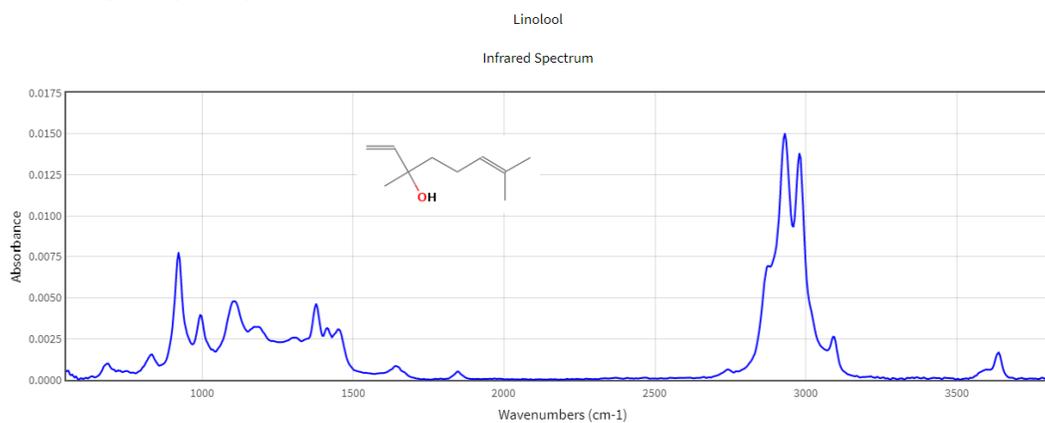
Acetato 2-metilfurano

Condiciones: SOLUTION (10% CCl₄ FOR 5000-1330, 10% CS₂ FOR 1330-625 CM-1)**Instrumento:** Not specified, most likely a prism, grating, or hybrid spectrometer.

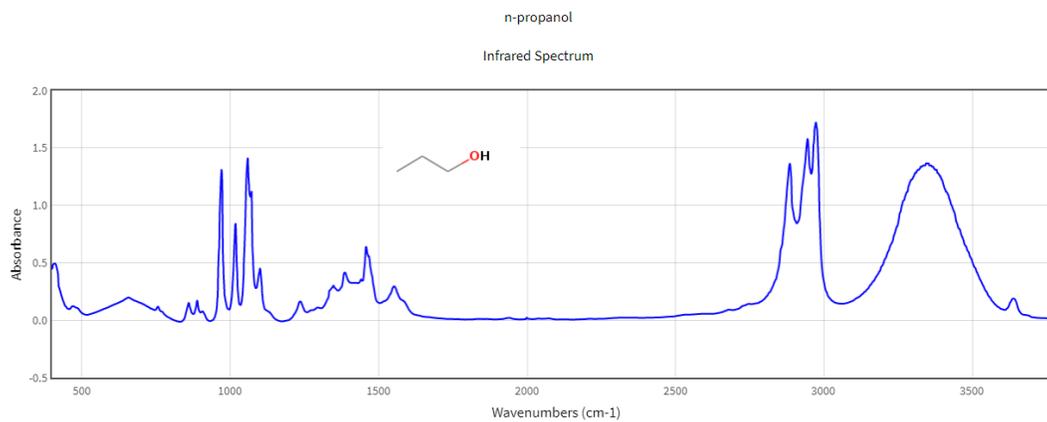
Farnesol

Condiciones: Sin información**Instrumento:** Sin información

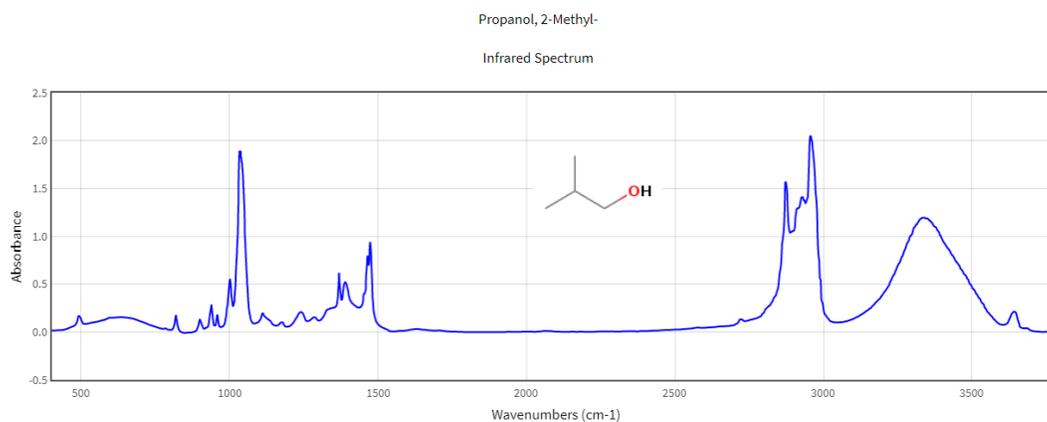
Linalol

Condiciones: Sin información**Instrumento:** Sin información

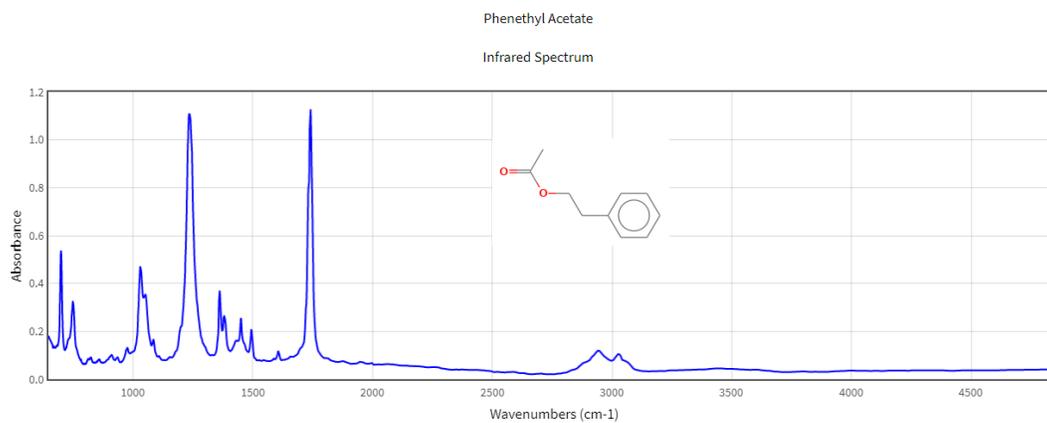
1-Propanol

Condiciones: SOLUTION (10% IN CCl4 FOR 3800-1330, 10% IN CS2 FOR 1330-400 CM-1)**Instrumento:** BECKMAN IR-9 (GRATING)

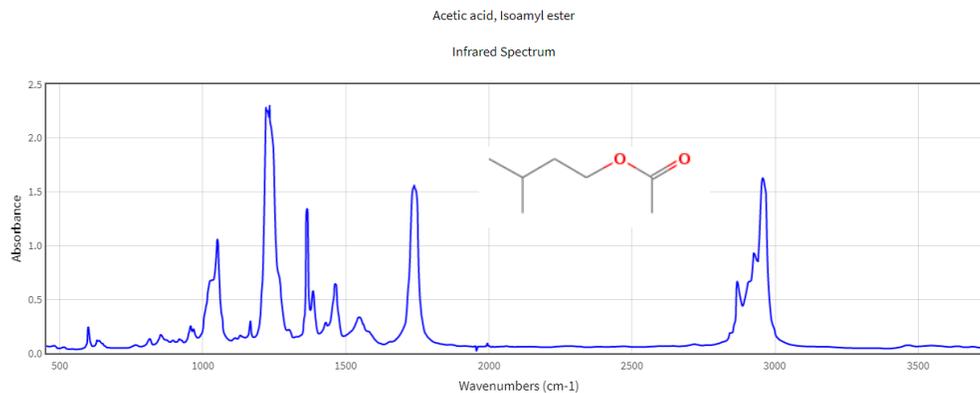
Iso-butanol

Condiciones: SOLUTION (10% IN CCl₄ FOR 3800-1340, 10% IN CS₂ FOR 1340-400 CM-1) VS. SOLVENT**Instrumento:** DOW KBr FOREPRISM-GRATING

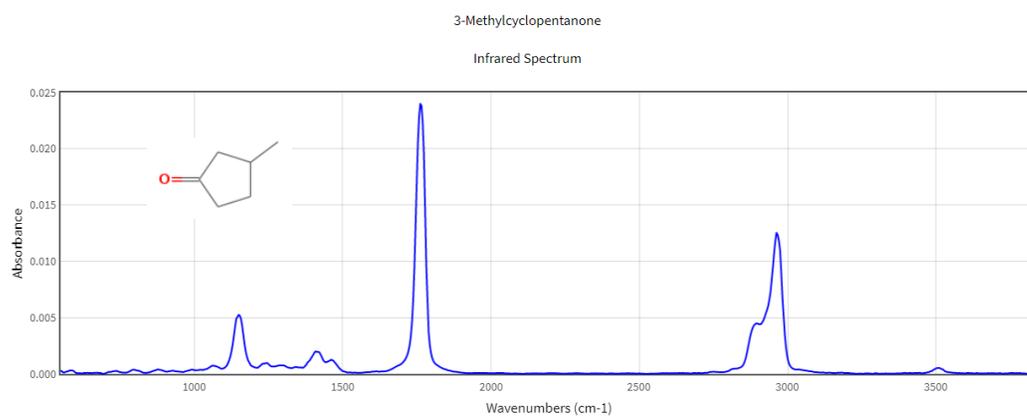
2-Etil Fenil Acetato

Condiciones: LIQUID**Instrumento:** PERKIN-ELMER

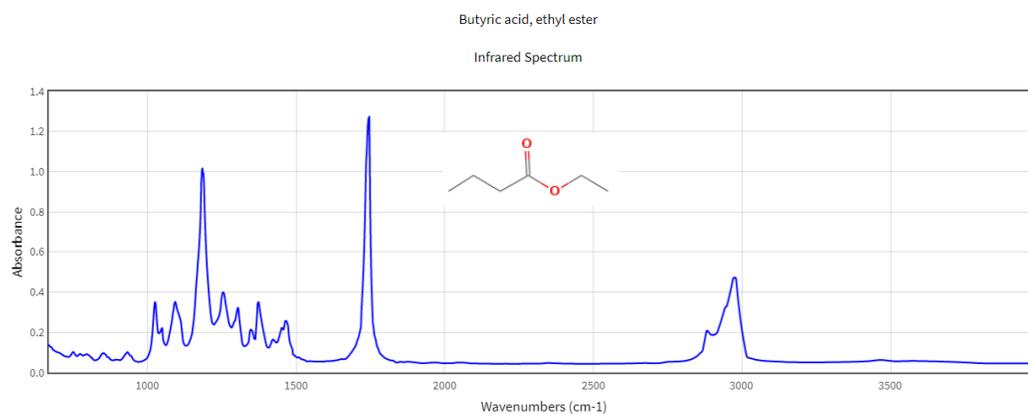
Acetato 3-metil 1-butanol

Condiciones: SOLUTION (10% CCl₄ FOR 3800-1330, 10% CS₂ FOR 1330-400 CM-1)**Instrumento:** DOW KBr FOREPRISM-GRATING

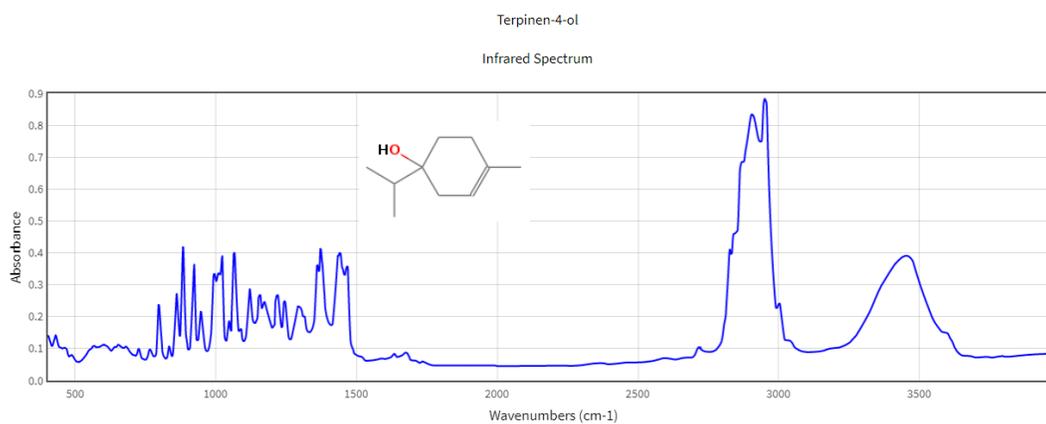
3-metilciclopentano

Condiciones: Sin referencias**Instrumento:** Sin referencias

Ácido butanoico, etil ester

Condiciones: LIQUID (NEAT)**Instrumento:** BECKMAN IR-4 (HYBRID PRISM-GRATING)

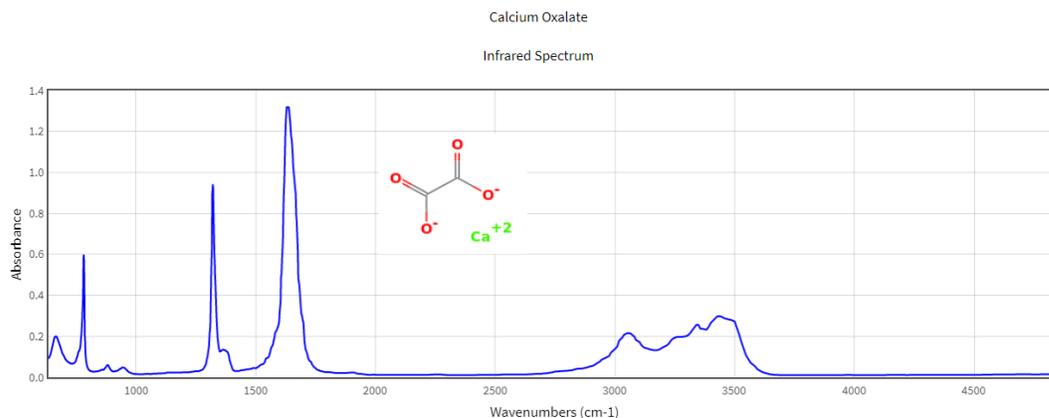
4-terpineol

Condiciones: LIQUID (NEAT)**Instrumento:** BECKMAN IR-12 (GRATING)

Oxalato de Calcio

Condiciones: Sin referencias

Instrumento: Sin referencias



11.5 Parámetros obtenidos para cada modelo construido en Matlab.

- Valores de Especificidad, Sensibilidad y Precisión para modelo de calibración, validación y predicción en un intervalo de 1550 a 400 cm^{-1} con la técnica de funciones de potencia.

fitting

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

validation

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00
not-assigned: 0.15

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

prediction on external samples

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

- Valores de Especificidad, Sensibilidad y Precisión para modelo de calibración, validación y predicción en un intervalo de 1550 a 850 cm^{-1} con la técnica de funciones de potencia.

fitting

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00
not-assigned: 0.17

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

prediction on external samples

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

validation

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00
not-assigned: 0.22

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

- Valores de Especificidad, Sensibilidad y Precisión para modelo de calibración y validación en un intervalo de 1550 a 850 cm^{-1} con la técnica de vecino más cercano (KNN).

fitting

error rate: 0.04
non-error rate: 0.96
accuracy: 0.96

class	Spec	Sens	Prec
1	0.91	1.00	0.94
2	1.00	0.91	1.00

validation

error rate: 0.04
non-error rate: 0.96
accuracy: 0.96

class	Spec	Sens	Prec
1	0.91	1.00	0.94
2	1.00	0.91	1.00

- Valores de Especificidad, Sensibilidad y Precisión para modelo de calibración y validación en un intervalo de 1550 a 400 cm^{-1} con la técnica de SIMCA.

fitting

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

validation

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

prediction on external samples

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

- Valores de Especificidad, Sensibilidad y Precisión para modelo de calibración y validación en un intervalo de 1550 a 850 cm^{-1} con la técnica de SIMCA.

fitting

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

validation

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00
not-assigned: 0.02

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00

prediction on external samples

error rate: 0.00
non-error rate: 1.00
accuracy: 1.00

class	Spec	Sens	Prec
1	1.00	1.00	1.00
2	1.00	1.00	1.00