



**UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
CUAUTILÁN**

**“Diferenciación de tequilas auténticos y destilados de agave
mediante métodos quimiométricos multivariables a partir de datos
de espectroscopía de infrarrojo medio”**

**T E S I S
QUE PARA OBTENER EL TÍTULO DE**

LICENCIADO EN QUÍMICA INDUSTRIAL

**P R E S E N T A
HUGO JIMÉNEZ FALCÓN**

ASESORES:

DRA. MARÍA GUADALUPE PÉREZ CABALLERO

DR. JOSÉ MANUEL ANDRADE GARDA

CUAUTILÁN IZCALLI, ESTADO DE MÉXICO, 2022



Universidad Nacional
Autónoma de México



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.

DEDICATORIAS

A mis padres:

A mis papás María Adelaida y Valentín porque siempre confiaron en mí y gracias a ellos puedo ser lo que soy hoy. De quienes recibí los mejores consejos de vida, los valores y el amor de familia.

A mis hermanas:

A mis hermanas Fernanda y Abril porque siempre estuvieron conmigo, siempre recibí el apoyo de su parte y por haberme enseñado importantes lecciones de vida.

A mi abuelo Esteban:

A mi abuelito Esteban porque desde niño siempre me inculcó la cultura del estudio y siempre mostró interés en que tuviera la mejor educación. Porque siempre estuvo conmigo en todo momento y, aunque no pueda estar presente en estos momentos, siempre recordaré con el más íntimo cariño.

AGRADECIMIENTOS

Primeramente agradezco a Dios por haberme dado la oportunidad de llegar hasta este punto de mi vida rodeado de tantas personas importantes.

A mis padres María Adelaida y Valentín por todo el esfuerzo que dedicaron a mi formación y crecimiento humano.

A mi asesora la Dra. Lupita Pérez Caballero por toda la paciencia que recibí de su parte, además de ser una persona que me enseñó muchísimo no solo en el ámbito profesional sino también de crecimiento humano.

A mi co-asesor el Dr. José Manuel Andrade Garda por todo el apoyo que nos brindó para la realización de este trabajo, por su valioso aporte de revisión y corrección del manuscrito y con quien aprendimos mucho durante el proceso de investigación.

A todos los amigos y compañeros de trabajo que hice durante mi paso por la FES-Cuautitlán con quienes pasé los mejores momentos, con quienes compartí buenas pláticas, fiestas, noches de desvelo y los problemas de la vida cotidiana. Disculparán que no los enumere a todos pero sé que cada uno sabe todo lo que pasamos juntos.

Finalmente agradezco a la Facultad de Estudios Superiores Cuautitlán-UNAM, por haberme abierto sus puertas y ofrecerme una formación excelente, donde conocí a muchos profesores que me motivaron día a día a seguir por el hermoso camino de la química.

INDICE

1.0 INTRODUCCIÓN.....	11
2.0 OBJETIVO GENERAL.....	14
3.0 OBJETIVOS PARTICULARES.....	14
4.0 MARCO TEÓRICO.....	15
4.1 El Tequila.....	15
4.2 Proceso de elaboración del tequila.....	17
4.3 Categorías y Clases de Tequila.....	22
4.4 Destilados y Licores de Agave.....	23
4.4.1 <i>Destilado de agave</i>	23
4.4.2 <i>Licor “de” o “con” agave</i>	23
4.5 ¿Por qué diferenciar tequilas auténticos de destilados/licores de agave?.	25
4.6 Espectrometría en la zona infrarroja media (FT-MIR)	25
4.7 Análisis Multivariante de datos o Quimiometría.....	27
4.8 Matriz de datos.....	28
4.9 Pretratamiento de datos.....	29
4.10 Escalados.....	30
4.10.1 <i>Centrado en la media</i>	30
4.10.2 <i>Autoescalado</i>	30
4.11 Técnicas de agrupación no supervisadas.....	31
4.11.1 <i>Análisis de Componentes Principales (PCA)</i>	31
4.11.2 <i>Análisis Cluster</i>	34
4.12 Precauciones con las técnicas de agrupamiento no supervisadas.....	37
4.13 Técnicas de clasificación supervisadas.....	37
4.13.1 <i>Curvas de Potencia</i>	38

4.13.2 Clasificación mediante vectores soporte, SVM (<i>Support Vector Machines</i>).....	39
5.0 DESARROLLO EXPERIMENTAL.	40
5.1 Muestras Analizadas.	40
5.2 Uso del Espectrómetro FT-IR.....	41
5.3 Pretratamiento de los datos.	42
5.4 Construcción de matrices.	43
5.5 Datan@.....	43
6.0 RESULTADOS OBTENIDOS AL TRATAR ESPECTROS ATR CON CORRECCIÓN DE LÍNEA BASE AUTOMÁTICA.....	45
6.1 Técnicas no supervisadas o de exploración de pautas.	46
6.1.1 Análisis de Componentes Principales (PCA).....	46
6.1.1.1 Modelos para la diferenciación de tequilas vs licores/destilados.....	46
6.1.1.2 Selección de intervalo de número de onda.....	46
6.1.1.3 Modelos para la diferenciación de licores/destilados vs tequila por clase.....	50
6.1.2 Análisis Cluster (CA)	53
6.1.2.1 Modelos de CA para la diferenciación de tequilas vs licores/destilados.	53
6.1.2.2 Modelos de CA para la diferenciación de licores/destilados vs tequila por clase	53
6.2 Técnicas supervisadas.....	56
6.2.1 Curvas de potencia.....	56
6.2.1.1 Modelos de CP para la diferenciación de licores/destilados vs tequila por clase.....	56
6.2.2 Clasificación mediante vectores soporte (<i>Support Vector Machines, SVM</i>)	62

<i>6.2.2.1 Modelos de SVM para la discriminación de tequilas vs licores/destilados.....</i>	<i>62</i>
<i>6.2.2.2 Modelos de SVM para la discriminación de licores/destilados vs tequila por clase.....</i>	<i>64</i>
7.0 CONCLUSIONES.....	69
8.0 REFERENCIAS.....	70
9.0 REFERENCIAS DE SOFTWARE.	71
10.0 REFERENCIAS DE FIGURAS.	71
11.0 ANEXOS.....	74

INDICE DE TABLAS Y FIGURAS

FIGURAS

Figura 01.- Distintas variedades de agave [García, 2016 a].....	11
Figura 02.-Distribución de agaves en la república mexicana [García, 2016 b].....	12
Figura 03.- Mayáhuel, diosa mexicana del agave (códice Borbónico) [Matadamas, 2016].....	15
Figura 04.- Producción total de Tequila y Tequila 100 % agave entre los años 1995-2018 [CRT, 2019 a].	16
Figura 05.- Geografía de la Denominación de Origen Tequila [CRT, 2019 b].	16
Figura 06.- Agave tequilana Weber variedad azul [García, 2016 c].	17
Figura 07.- Jima del agave [García, 2016 d].	18
Figura 08.- Horno de mampostería [García, 2016 e].	19
Figura 09.- Tahona [García, 2016 f].	19
Figura 10.- Tanques de fermentación [García, 2016 g].	20
Figura 11.- Alambiques [García, 2016 h].	21
Figura 12.- Barricas [García, 2016 i].	22
Figura 13.- Representación de una Matriz de datos en una hoja de cálculo de Excel® [Excel®, 2013].	29
Figura 14.- a) Sin Escalado, b) Centrado en la media, c) Autoescalado [Otto, 2017 a].....	31
Figura 15.- Representación gráfica de la matriz X en PCA [Otto, 2017 b].	32
Figura 16.- Representación gráfica de los PCs [Otto, 2017 c].	32
Figura 17.-Unión de Clusters con distancia Single Linkage.	35
Figura 18.- Forma típica de un dendrograma.	36
Figura 19.- Presentación de las muestras utilizadas en este trabajo.	40
Figura 20.- Efecto de los diferentes pretratamientos en los datos FIT-MIR.	42
Figura 21.- Formato de una matriz en Excel®.....	43
Figura 22.- Vista previa de Datan® con una matriz de datos.	44

Figura 23.- Gráfico de scores para el conjunto de licores/destilados y tequilas en el intervalo: 400-4000 cm^{-1} , Sin escalado. PC1: 99.9571 %, PC2: 0.0306 %, PC3: 0.0079 %.....	46
Figura 24.- Loadings asociados a los primeros componentes principales en el intervalo de 400 - 4000 cm^{-1}	47
Figura 25.- Perfiles de los espectros FT-MIR con diferentes escalados, de arriba hacia abajo: sin escalado, centrado en la media y autoescalado. Intervalo optimo de trabajo resaltado en rosa.	48
Figura 26.- Diagrama de scores en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Sin escalado. PC1: 99.8138 %, PC2: 0.1561 %, PC3: 0.0246 %.....	49
Figura 27.- Diagrama de scores en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media. PC1: 90.8301 %, PC2: 6.9695 %, PC3: 1.6127 %.....	49
Figura 28.- Diagrama de scores en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 61.8607 %, PC2: 21.8087 %, PC3: 11.1533 %.....	50
Figura 29.- Diagrama de scores destilados/licores vs. tequilas blancos en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 70.7163 %, PC2: 88.258 %, PC3: 96.9365 %.....	51
Figura 30.- Diagrama de scores destilados/licores vs. tequilas jóvenes en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 63.6495 %, PC2: 85.6239 %, PC3: 95.4386 %.....	51
Figura 31.- Diagrama de scores destilados/licores vs. tequilas reposados en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 62.0827 %, PC2: 86.0065 %, PC3: 95.2580 %.....	52
Figura 32.- Diagrama de scores destilados/licores vs. tequilas añejos/extra añejos en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 62.6431 %, PC2: 84.8454 %, PC3: 95.0718 %.....	52
Figura 33.- Dendrograma para el conjunto de destilados/licores y tequilas, algoritmo de Ward y distancia Euclidea cuadrada. Autoescalado, intervalos: 980-1800 cm^{-1} y 2800-3200 cm^{-1}	53
Figura 34.- Dendrograma para el conjunto de destilados/licores y tequilas blancos, algoritmo de Ward y distancia Euclidea cuadrada. Autoescalado, intervalos: 980-1800 cm^{-1} y 2800-3200 cm^{-1}	54
Figura 35.- Dendrograma para el conjunto de destilados/licores y tequilas jóvenes, algoritmo de Ward y distancia Euclidea cuadrada. Autoescalado, intervalos: 980-1800 cm^{-1} y 2800-3200 cm^{-1}	54

Figura 36.- Dendrograma para el conjunto de destilados/licores y tequilas reposados, algoritmo de Ward y distancia Euclidea cuadrada. Autoescalado, intervalos: 980-1800 cm^{-1} y 2800-3200 cm^{-1}	55
Figura 37.- Dendrograma para el conjunto de destilados/licores y tequilas añejos/extra añejos, algoritmo de Ward y distancia Euclidea cuadrada. Autoescalado, intervalos: 980-1800 cm^{-1} y 2800-3200 cm^{-1}	55
Figura 38.- Curvas de potencia para destilados/licores vs. tequilas blancos en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media.	57
Figura 39.- Curvas de potencia para destilados/licores vs. tequilas jóvenes en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media.	58
Figura 40.- Curvas de potencia para destilados/licores vs. tequilas reposados en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media.	59
Figura 41.- Curvas de potencia para destilados/licores vs. tequilas añejos/extra añejos en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media.....	60
Figura 42.- Análisis óptimo para SVM considerando Destilados/licores y tequilas en todas sus clases en los intervalos 980-1800 cm^{-1} , 2800-3200 cm^{-1} , Kernel lineal, C=50, 10000 iteraciones.	63
Figura 43.- Grafico SVM para destilados/licores vs. tequilas blancos, Centrado en la media en los intervalos 980-1800 cm^{-1} , 2800-3200 cm^{-1} , Kernel lineal, C=50, 10000 iteraciones..	64
Figura 44.- Grafico SVM para destilados/licores vs. tequilas jóvenes, Centrado en la media en los intervalos 980-1800 cm^{-1} , 2800-3200 cm^{-1} , Kernel lineal, C=50, 10000 iteraciones....	65
Figura 45.- Grafico SVM para destilados/licores vs. tequilas reposados, Centrado en la media en los intervalos 980-1800 cm^{-1} , 2800-3200 cm^{-1} , Kernel lineal, C=50, 10000 iteraciones....	66
Figura 46.- Grafico SVM para destilados/licores vs. tequilas añejos/extra añejos, Centrado en la media en los intervalos 980-1800 cm^{-1} , 2800-3200 cm^{-1} , Kernel lineal, C=50, 10000 iteraciones.	67

TABLAS

Tabla 1. Distancias comunes en Cluster y sus ecuaciones.	34
Tabla 2. Probabilidades de pertenencia de las muestras de validación para destilados/licores vs tequilas blancos.....	57
Tabla 3. Probabilidades de pertenencia de las muestras de validación para destilados/licores vs tequilas jóvenes.....	58
Tabla 4. Probabilidades de pertenencia de las muestras de validación para destilados/licores vs tequilas reposados.	59
Tabla 5. Probabilidades de pertenencia de las muestras de validación para destilados/licores vs tequilas añejos/extra añejos.	60
Tabla 6. Análisis de capacidad de predicción para CP.	61
Tabla 7. Selección de muestras de validación.	62
Tabla 8. Asignación de las muestras de validación para destilados/licores vs tequilas vía SVM.....	63
Tabla 9. Asignación de las muestras de validación para destilados/licores vs tequilas blancos vía SVM.....	64
Tabla 10. Asignación de las muestras de validación para destilados/licores vs tequilas jóvenes vía SVM.....	65
Tabla 11. Asignación de las muestras de validación para destilados/licores vs tequilas reposados vía SVM.....	66
Tabla 12. Asignación de las muestras de validación para destilados/licores vs tequilas añejos/extra añejos vía SVM.....	67
Tabla 13. Análisis de capacidades de reconocimiento y predicción para SVM.....	68

1.0 INTRODUCCIÓN

México es un país con gran cultura y tradiciones, algunas de las cuales se remontan a la época prehispánica. Las culturas prehispánicas aprendieron a dominar su entorno, aprovechando los recursos que la naturaleza les proporcionó, esto incluyó la flora y fauna, entre ellos el agave.

El agave es una planta endémica de América. Se conocen aproximadamente 210 especies de las cuales 150 se encuentran en México. Eso lo ubica como el país con mayor número de especies, aproximadamente un 75 % de las especies totales [García, 2007].



Figura 01.- Distintas variedades de agave [García, 2016 a].

El agave o maguey –como se le conoce en diversas partes del país- esta a su vez distribuido ampliamente, teniendo presencia en más del 75 % del territorio nacional [García, 2007], por lo cual esta planta es tan importante y conocida por prácticamente todos los mexicanos.

Antes de la llegada de Colón diversas especies de agave se aprovechaban para producir fibras naturales –henequén, ixtle- así como forraje para el ganado y bebidas alcohólicas fermentadas como el pulque [Parsons, 2012]. Con la llegada de los europeos al continente americano llegó una nueva técnica, la destilación, la cual permitió la producción de bebidas alcohólicas destiladas derivadas de ciertas especies de agave. A finales del siglo XVIII se otorgó la primera concesión virreinal para la producción de tequila en México. José Cuervo (fundador de “Casa Cuervo”) fue el acreedor de la concesión. El tequila representa el encuentro de dos mundos, incorpora al agave oriundo de América y las técnicas europeas de industrialización.

El sotol, la bacanora, el mezcal y el tequila son ejemplos de bebidas alcohólicas producidas en México a partir de plantas de agave como materia prima principal. Estas bebidas cuentan todas con Denominación de Origen [García, 2012]. Una denominación de origen implica que cierto producto presenta ciertas cualidades específicas debido a factores geográficos, ambientales, flora y/o fauna exclusiva de cierta región. Esto trae como consecuencia que se restrinja la elaboración de un producto a una región, con condiciones ambientales e insumos específicos. De manera particular el tequila cuenta con un organismo llamado Consejo Regulador del Tequila (CRT) que se encarga de verificar que este sea producido de acuerdo a las especificaciones establecidas en la NOM-006-SCFI-2012 [Norma Oficial Mexicana 006, 2012]. Casos análogos aplican para las demás bebidas con denominación de origen y sus normas regulatorias, Sotol: NOM-159-SCFI-2004 [Norma Oficial Mexicana 159, 2004], Bacanora: NOM-168-SCFI-2004, [Norma Oficial Mexicana 168, 2004], Mezcal: NOM-070-SCFI-2016, [Norma Oficial Mexicana 070, 2016].

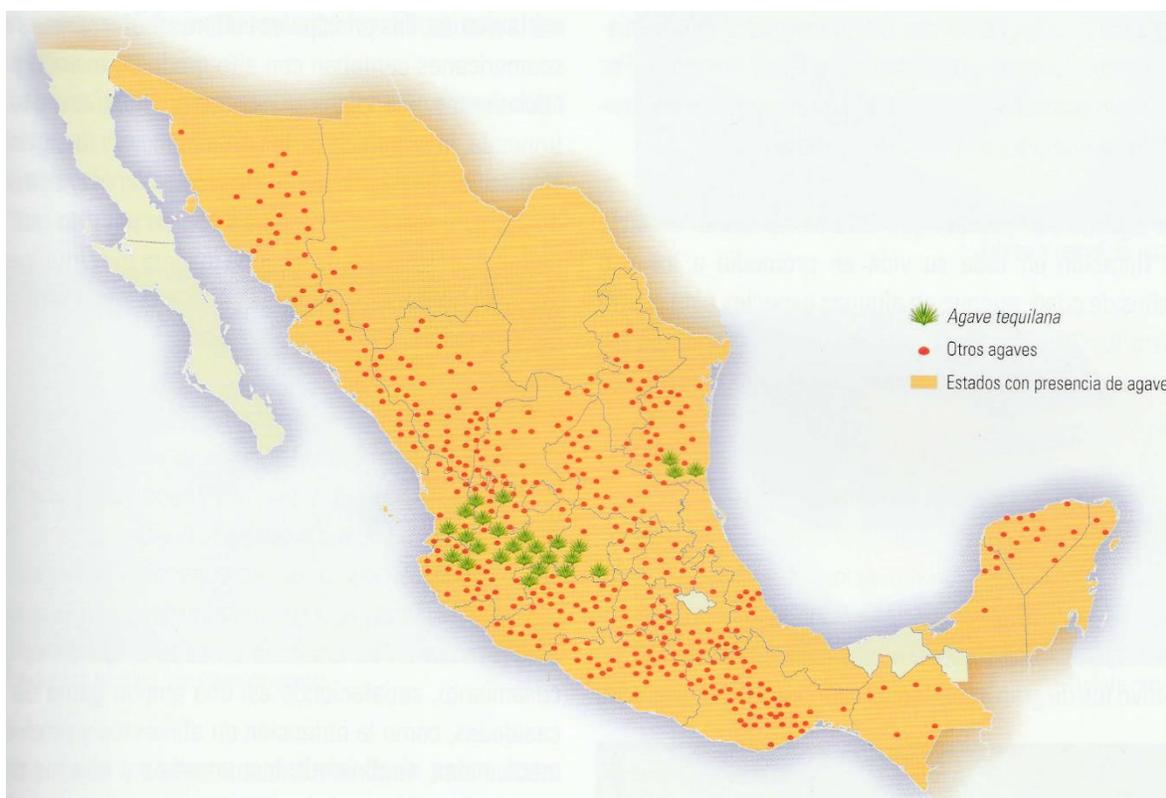


Figura 02.-Distribución de agaves en la república mexicana [García, 2016 b].

El tequila es, sin lugar a dudas, la bebida alcohólica mexicana más conocida a nivel mundial. El 9 de diciembre de 1974 se publicó la Declaratoria de Protección de la Denominación de Origen del Tequila en el Diario Oficial de la Federación. Aunque es una bebida que ha sido producida en el país por muchos años, fue a partir de que obtuviera la denominación de origen cuando comenzó a volverse cada vez más popular. En años recientes se ha visto que el tequila se vuelve cada vez más importante en la vida social mexicana, impulsado por la

mercadotecnia, productos como el “tequila cristalino”, así como un cada vez más creciente sentimiento de nacionalidad entre la población.

La gran variedad de especies de agave han facilitado la producción de diversas bebidas alcohólicas. Existen en el mercado los denominados destilados de agave, además de los licores de agave. Estas bebidas no cuentan con ninguna o casi ninguna regulación por lo que el gobierno mexicano se vio en la necesidad de proponer una nueva norma que incluya a estas bebidas.

En el año 2006 se decidió tomar medidas para comenzar el proceso regulatorio de los destilados/licores de agave por medio del establecimiento de una NOM de emergencia. Actualmente se encuentra aún en propuesta por la Comisión Federal de Mejora Regulatoria (COFEMER) manteniéndose como un “proyecto de NOM”, esta es el PROY-NOM-186-SCFI-2012, [Proyecto de Norma Oficial Mexicana 186, 2012]. En este proyecto de norma oficial mexicana se dan las descripciones de las diversas bebidas alcohólicas producidas a partir de agave, haciendo hincapié en que no se incluyen a bebidas que ya estén reguladas por alguna NOM particular, o que cuenten con denominación de origen.

El proceso de destilación permite separar componentes de una mezcla de acuerdo a sus puntos de ebullición. Al aplicarse esta técnica a bebidas alcohólicas como el tequila se debe tener el cuidado de evitar mantener las fracciones conocidas como “cabeza” y “cola” del destilado, ya que es en estas fracciones donde se recuperan compuestos como el metanol y algunos de los alcoholes superiores, los cuales son mucho más tóxicos para el cuerpo humano. En el tequila comercial es mucho menos probable encontrar residuos de estos “alcoholes tóxicos” que en los denominados destilados de agave por las diferencias en sus marcos regulatorios. No obstante, resulta de suma importancia el poderlos diferenciar adecuadamente.

En el presente trabajo se expone una metodología para la diferenciación entre tequilas y destilados de agave por medio de técnicas quimiométricas multivariantes, a partir de mediciones experimentales de espectros en la zona infrarroja de diferentes tequilas y destilados de agave.

2.0 OBJETIVO GENERAL:

Obtener modelos óptimos por medio de técnicas quimiométricas multivariantes haciendo uso de datos de espectrofotometría de infrarrojo medio que permitan lograr la diferenciación de tequilas auténticos de destilados/licores de agave.

3.0 OBJETIVOS PARTICULARES:

1. Diferenciar tequilas auténticos de destilados/licores de agave por medio de dos técnicas no supervisadas: Análisis de Componentes Principales (PCA) y Análisis Cluster (CA).
2. Lograr la discriminación de tequilas auténticos de los destilados/licores de agave por medio de dos técnicas supervisadas: *Support Vector Machines* (SVM) y Curvas de Potencia (PC).

4.0 MARCO TEÓRICO

Leyenda prehispánica sobre el origen del agave



Figura 03.- Mayáhuil, diosa mexicana del agave (códice Borbónico) [Matadamas, 2016].

“Mayáhuil era una hermosa joven que vivía con su abuela en las estrellas. En una ocasión, Quetzalcóatl la convenció para que bajase a la tierra para amarse convertidos en las ramas de un árbol bifurcado. Pero cuando su abuela se despertó y no vio a Mayáhuil decidió buscar a su nieta. La abuela, descubriendo a su nieta convertida en rama, la despedazó. Sin embargo la rama en que se había convertido Quetzalcóatl permaneció intacta. Quetzalcóatl tomó los restos de la joven virgen y los enterró. De ello brotó la planta del maguey. Así, tras su muerte, Mayáhuil se convirtió en diosa.”

4.1 El Tequila

El tequila es una bebida alcohólica de origen mexicano. Técnicamente es un tipo de mezcal. Su origen se remonta al siglo XVI aproximadamente. De acuerdo a la NOM-006-SCFI-2012 el tequila es:

“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, derivados de las cabezas de *Agave tequilana* Weber variedad azul, previa o posteriormente hidrolizadas o cocidas, y sometidos a fermentación alcohólica con levaduras, cultivadas o no, siendo susceptibles los mostos de ser enriquecidos y mezclados conjuntamente en la formulación con otros azúcares hasta en una proporción no mayor de 49 % de azúcares reductores totales expresados en unidades de masa, en los términos establecidos en la presente norma y en la inteligencia que no están permitidas las mezclas en frío. El Tequila es un líquido que puede tener color, cuando sea madurado, abocado, o añadido de un color específico. El Tequila puede ser añadido de edulcorantes, colorantes, aromatizantes y/o saborizantes permitidos por la Secretaría de Salud, con objeto de proporcionar o intensificar su color, aroma y/o sabor”.

Esta bebida es con seguridad la bebida alcohólica mexicana más conocida a nivel mundial, y cada año su popularidad aumenta entre la población nacional y extranjera. De acuerdo a estadísticos del CRT la producción de tequila aumenta con el pasar de los años. Esto puede verse claramente en la figura 4.

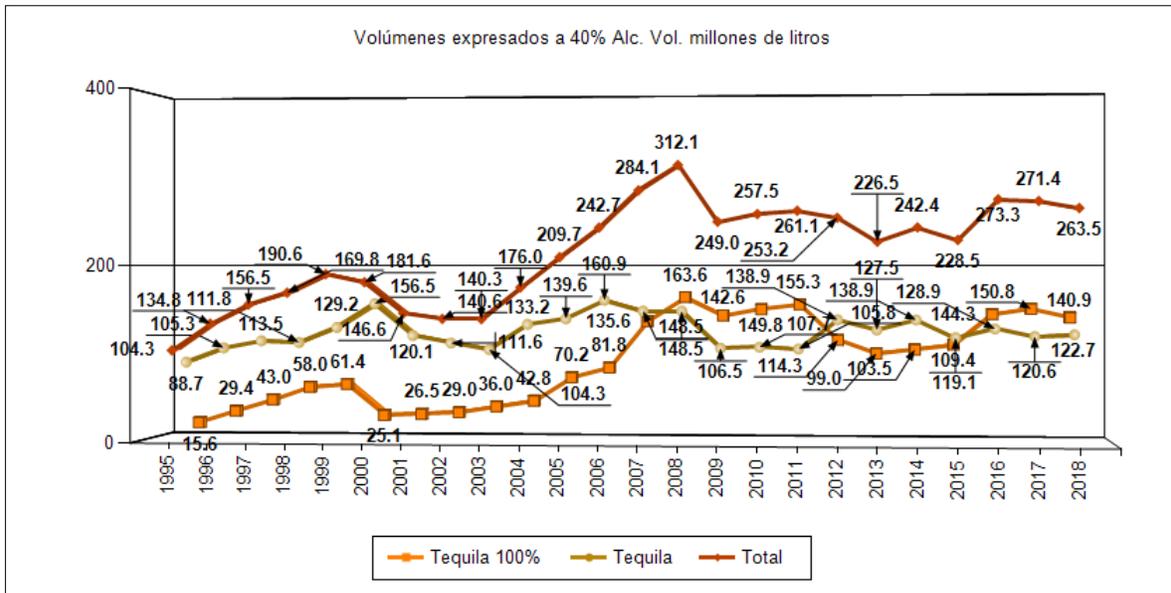


Figura 04.- Producción total de Tequila y Tequila 100 % agave entre los años 1995-2018 [CRT, 2019 a].

La denominación de origen tequila comprende el estado de Jalisco, y algunos municipios de Tamaulipas, Nayarit, Guanajuato y Michoacán. La figura 5 muestra los municipios de los estados productores de tequila.



Figura 05.- Geografía de la Denominación de Origen Tequila [CRT, 2019 b].

4.2 Proceso de elaboración del tequila

La elaboración del tequila es un proceso largo que requiere de paciencia y un extremo cuidado. Todo comienza con la plantación del *Agave tequilana* Weber (figura 6) el cual requiere aproximadamente 7 años para poder estar listo para la producción de tequila. Después de estos 7 años el agave tiene la mayor cantidad de carbohidratos hidrolizables necesarios para la fermentación alcohólica.



Figura 06.- *Agave tequilana* Weber variedad azul [García, 2016 c].

El siguiente paso se conoce como “Jima” y consiste en que el “Jimador” por medio de una herramienta especial llamada “Coa” coseche el agave (figura 7). La coa no es más que una navaja con un mango largo que le permite al jimador retirar las hojas (pencas) del agave para quedarse únicamente con el corazón del agave, comúnmente llamado “piña”. Es importante remarcar que aunque actualmente la industria tequilera realiza muchos de sus procesos por medio de máquinas automatizadas sigue manteniendo algunas partes de forma artesanal.



Figura 07.- Jima del agave [García, 2016 d].

El siguiente paso es el transporte de las piñas desde el campo de cultivo hasta las instalaciones de la planta tequilera. Tanto el campo de cultivo como la planta tequilera deben estar ubicados en los territorios que la “denominación de origen tequila” comprende.

Las piñas frescas son introducidas en hornos de mampostería (figura 8) del tamaño de habitaciones grandes donde son cocidos “al vapor” a una temperatura de entre 100-110 °C. La inulina es un polímero de fructosa y glucosa y es el principal carbohidrato presente en el agave. El proceso térmico permite la hidrólisis de la inulina para obtener carbohidratos más sencillos necesarios para la fermentación. La hidrólisis juega un papel muy importante en el resultado final. Si no se regula la temperatura adecuadamente, parte de los carbohidratos puede convertirse en caramelo (no fermentable), además de que el proceso de caramelización trae consigo la producción de furfural (inhibidores de la fermentación). Por otra parte si no se controla el tiempo de hidrólisis puede generarse una producción excesiva de metanol proveniente de la desmetilación de las pectinas presentes en la fibra. Este metanol crearía mezclas azeotrópicas con el etanol y el agua y daría como resultado un tequila de baja calidad.

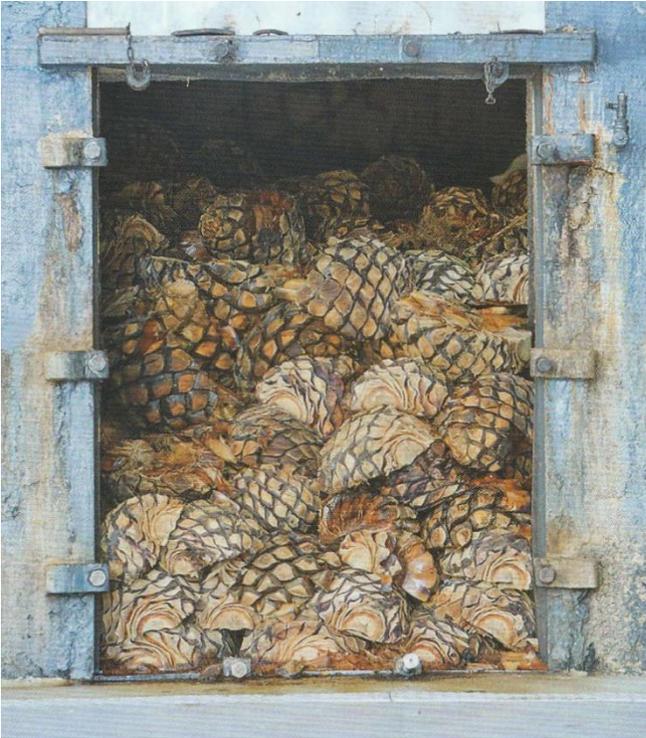


Figura 08.- Horno de mampostería [García, 2016 e].

Después de la hidrólisis se realiza la extracción de los jugos de las piñas cocidas. Esto se realiza por medio de extrusión mecánica por medio de rodillos. En algunas casas tequileras esto se realiza aun en tahonas (figura 9) aunque cada vez es menos común su uso y menos aún el uso de animales para impulsarlas.



Figura 09.- Tahona [García, 2016 f].

Aquí se llega a un punto crítico en la producción del tequila; es necesario decidir si se producirá un tequila 100 % de agave o un tequila mixto. Si se opta por el tequila 100 % agave se usará únicamente el jugo del agave cocido, si por otra parte se decide producir un tequila mixto este jugo se combinará con no más de 49 % de otras fuentes de carbohidratos externas. Ajustando la mezcla, el pH y temperatura se obtiene el mosto fresco listo para la fermentación. Este proceso se conoce como formulación y su importancia radica en que la elección de la categoría de tequila determinará el costo económico del producto final, teniéndose como lineamiento que un tequila 100 % agave es más costoso que un tequila mixto.

El siguiente paso consiste en pasar el mosto fresco a los tanques de fermentación (figura 10). Se regulan la temperatura y el pH, además se adicionan las levaduras que se encargarán de realizar la fermentación alcohólica; el proceso general puede resumirse fundamentalmente en la siguiente reacción bioquímica:



Las levaduras usadas en la fermentación suelen ser exclusivas de cada destilería, pudiendo ser además levaduras comerciales que comúnmente están compuestas por levaduras de vino, pan y una mezcla comercial conocida como levadura de tequila. Existe también la posibilidad de que las destilerías más exigentes cuenten con su propia levadura aislada de manera natural de los agaves locales [Williams, 2015]. La fermentación dura un aproximado de 72 horas.



Figura 10.- Tanques de fermentación [García, 2016 g].

A continuación el mosto fermentado es enviado a los alambiques de destilación (figura 11). La destilación está dividida a su vez en dos fases, la primera de ellas se conoce como destrozamiento o agotamiento y consiste en la separación de las vinanzas –mezcla de levadura, sales, sólidos, alcoholes secundarios y superiores, agua- de los componentes

volátiles de la mezcla. La segunda fase denominada rectificación o enriquecimiento consiste en una destilación que corrige aquello que no se logró durante el destrozamiento.



Figura 11.- Alambiques [García, 2016 h].

Una buena destilación será aquella que permita separar la parte “tóxica” de la mezcla de mostos fermentados y que, además, no sea tan rigurosa para separar aquellos componentes volátiles provenientes del agave que son responsables de las características sensoriales del tequila [Williams, 2015]. El tequila obtenido tiene un contenido alcohólico demasiado alto por lo que se adiciona agua de dilución para llegar a un contenido alcohólico de 35-55 % Alc. Vol. [Norma Oficial Mexicana 006, 2012], siendo el rango de 38-40 % Alc. Vol. el más común en los tequilas comerciales.

Finalmente, el tequila es llevado a las barricas de maduración (figura 12); éstas deben estar fabricadas con madera de roble o encino. Es en las barricas donde el tequila adquiere notas características de cada una de las clases de tequila existentes. Finalizada la maduración el tequila es embotellado y comercializado.



Figura 12.- Barricas [García, 2016 i].

4.3 Categorías y Clases de Tequila

El tequila puede clasificarse dependiendo del porcentaje de azúcares provenientes del agave en su elaboración, esto da como resultado una división de dos categorías: Tequila 100 % agave y Tequila mixto; llamado simplemente Tequila.

El tequila 100 % agave es aquel en el que se usan exclusivamente azúcares provenientes del *Agave tequilana* Weber variedad azul. Un tequila mixto es aquel en el que como mínimo se usa un 51 % de azúcares provenientes del *Agave tequilana* Weber variedad azul, el 49 % restante puede provenir de fuentes externas.

También es posible clasificar al tequila en base al tiempo de maduración que pasó en barrica. Surgen aquí las clases de tequila: Blanco o Plata, Joven u Oro, Reposado, Añejo y Extra Añejo.

El tequila blanco es aquel obtenido inmediatamente de la destilación pudiendo pasar en barricas un tiempo menor a dos meses. El tequila joven es una mezcla de tequila blanco con tequila reposado, añejo o extra añejo; también se considera tequila joven a un tequila blanco que haya sido sujeto de abocamiento. Tequila reposado es el que ha pasado por lo menos dos meses en barrica y menos de un año. Un tequila añejo debe madurar mínimo un

año y menos de tres en una barrica que no sobrepase los 600 litros de volumen. Finalmente un tequila Extra Añejo tendrá por lo menos 3 años de maduración en barrica de no más de 600 litros de volumen.

4.4 Destilados y Licores de Agave

La definición de un destilado de agave y de un licor de agave es por mucho una tarea complicada. Es importante aclarar que la simple referencia “Agave” en la denominación de estos productos genera confusión en gran parte de la población. Esto se debe principalmente a que la industria del tequila y los mezcales en general utilizan la palabra Agave en sus productos. Es entonces Agave la palabra clave en el dilema.

La mejor definición que se puede tener hasta el momento de la identidad de un destilado de agave y de un licor de agave la proporciona el Proyecto de Norma Oficial Mexicana PROY-NOM-186-SCFI-2012. Este los define de la siguiente manera:

4.4.1 Destilado de agave:

Bebida alcohólica obtenida por destilación y rectificación parcial de mostos fermentados, derivados de la molienda de las cabezas de agave, previa o posteriormente hidrolizadas o cocidas, y sometidos a fermentación alcohólica, siendo susceptibles los mostos de ser enriquecidos por otros azúcares hasta en una proporción no mayor de 49 % de azúcares reductores totales expresados en unidades de masa, no estando permitidas las mezclas en frío. El destilado de agave es un líquido que, de acuerdo a su tipo, es incoloro o amarillento cuando se aboque sin madurarlo.

El Destilado de Agave es susceptible de ser adicionado de algún sabor o ingrediente permitido por la Secretaría de Salud, su contenido alcohólico es de 25 % a menos de 35 % Alc. Vol. [Proyecto de Norma Oficial Mexicana 186, 2012].

Los destilados de agave pueden clasificarse en dos clases: Destilado de agave blanco y Destilado de agave abocado. La única diferencia entre las dos clases de Destilado de agave consiste en que el blanco es un producto incoloro y el abocado tiene coloración debida a la incorporación de aditivos permitidos por la secretaria de salud.

4.4.2 Licor “de” o “con” agave:

Son los productos elaborados a partir de Aguardientes de agave, Destilados de agave y en su caso otros tales como, espíritu neutro, alcohol de calidad o alcohol común o mezcla de ellos; con un contenido no menor de 1 % (m/v) de azúcares y agua; aromatizados y saborizados con procedimientos específicos y que pueden adicionarse de ingredientes así como de aditivos y coadyuvantes permitidos por la Secretaría de Salud; su contenido alcohólico es de 13.5 a 55 % Alc. Vol. [Proyecto de Norma Oficial Mexicana 186, 2012].

Aquí existe un punto crítico en la definición de los licores de agave. No es extraño que el subtítulo haga hincapié en las palabras “de” y “con” ya que la diferencia en estos nexos implica un cambio radical en la composición de un licor.

Un Licor de Destilado de Agave o de Aguardiente de Agave es aquel en el que la composición del licor contiene por lo menos 51 % de aguardiente o destilado de agave.

Un Licor con Destilado de Agave o con Aguardiente de Agave es aquel en el que la composición del licor contiene no menos de 1 % de aguardiente o destilado de agave.

Los licores de/con agave pueden clasificarse de la siguiente manera:

- a) Extra-seco: cuando el licor contenga de 1.0 % a 1.9 % de azúcares (m/v).
- b) Seco: cuando el licor contenga de 2.0 % a 10 % de azúcares (m/v).
- c) Dulce: cuando el licor contenga más de 10.1 % de azúcares (m/v).
- d) Escarchado: cuando el producto se encuentre sobresaturado de azúcares y en el cual el exceso de éstos haya cristalizado dentro del licor.
- e) Aperitivo: cuando el producto contenga algunos principios amargos derivados de sustancias vegetales.

Teniendo en cuenta las definiciones anteriores y las clasificaciones de los destilados y licores de agave es importante resumir y rescatar los puntos más importantes recordando la definición de los Tequilas.

1. No existen categorías para los destilados y licores de agave, simplemente se requiere que cuente con mínimo 51 % para destilados y 1 % para licores, de azúcares provenientes de agave en su producción.
2. Las clases de destilados de agave solo hacen referencia al proceso de abocamiento, no hay restricciones en cuanto a maduración ya que para estas bebidas no es necesario ninguna clase de reposo.
3. Los licores de agave no requieren de la planta de agave para su elaboración ya que éstos se producen a partir de otras bebidas que si lo demandan en su producción.
4. La elaboración de los licores de agave permite el uso de alcohol potable o similares siempre y cuando el producto tenga no menos de 1 % de azúcares y agua.
5. Únicamente se requiere el uso de agave sin tomar en cuenta ninguna especie en particular.

4.5 ¿Por qué diferenciar tequilas auténticos de destilados/licores de agave?

La principal razón es, sin duda, por el bien de los consumidores y del público en general. También para regular la comercialización de productos que en ocasiones hacen uso de publicidad engañosa.

Es imperativo diferenciar los Tequilas de los Destilados/Licores de agave ya que los segundos representan una competencia desleal contra los primeros. La justificación de lo anterior es lo siguiente:

1. Sus procesos de elaboración son distintos, tanto en materias primas, tiempos de producción como regulaciones normativas.
2. Los Destilados/Licores de agave no cuentan con un organismo regulatorio específico para ellos, a diferencia del Tequila que cuenta con el CRT. Esto implica que el Tequila tiene una mayor regulación.
3. Se ha visto que Destilados/Licores de agave hacen uso de publicidad engañosa que infunde una idea errónea en la población respecto al origen e identidad de estos productos.
4. El tequila es una bebida reconocida que cuenta con Denominación de Origen que debe protegerse de bebidas similares que no obedecen a las restricciones que una Denominación de Origen representa.

4.6 Espectrometría en la zona infrarroja media (FT-MIR)

La luz infrarroja es un tipo de radiación electromagnética ubicada por debajo de la región de la luz visible, de longitud de onda más larga, de menor frecuencia y menor energía que esta. Normalmente es percibida en forma de calor.

Cuando uno piensa en un enlace entre dos átomos es común que muchos imaginen un enlace estático y sólido. Sin embargo esto no sucede así, los enlaces siempre se encuentran en movimiento, estos movimientos se pueden clasificar en estiramientos y flexiones. El resultado de estos movimientos es la vibración en el enlace y por consiguiente en la molécula que lo contiene. Cada enlace tiene su modo específico de vibración y lo realiza a una frecuencia específica.

La espectrometría de infrarrojo medio es una técnica que utiliza la radiación infrarroja en la zona media del rango IR y su interacción con una muestra sólida, líquida o gaseosa. El resultado de esta interacción es el espectro infrarrojo de la sustancia o mezcla de sustancias.

El espectro de infrarrojo es la representación gráfica de las absorciones que sufre una muestra cuando es irradiada con radiación policromática. Cuando un enlace es irradiado con una frecuencia que coincida con uno de sus modos vibracionales el enlace absorbe energía y esto provoca que se estire o flexione un poco más [Yurkanis, 2008]. El espectro completo de IR es una propiedad altamente característica de un compuesto [Levine, 1991].

Sin embargo, no todos los enlaces absorben energía en la región del infrarrojo, para que esto suceda es necesario que exista un cambio en el momento dipolar de la molécula que los contiene [Christian, 2009].

Las estructuras orgánicas cuentan con diversos grupos funcionales en ellas, estos grupos funcionales tienen la propiedad de absorber en la región del infrarrojo. Y lo que es más importante aún, lo hacen a números de onda específicos e intensidades características. La espectrometría infrarroja es muy utilizada en química orgánica cuando de elucidación estructural se trata.

Un espectro de infrarrojo se divide principalmente en dos zonas importantes:

1. Región de grupos funcionales ($4000-1400\text{ cm}^{-1}$): Aquí se encuentran las principales y más intensas (en general) señales debidas a los grupos funcionales.
2. Región dactiloscópica ($1400-600\text{ cm}^{-1}$): Esta región es característica de cada compuesto, también es conocida como “región de huellas dactilares” y permite identificar un compuesto si se compara con un espectro IR de una muestra estándar. Pero su interpretación detallada no suele ser sencilla.

Un tequila es una mezcla muy compleja que va más allá de alcohol y agua. Técnicas analíticas como la cromatografía de gases acoplada a espectrometría de masas han permitido detectar e identificar más de 240 componentes volátiles en los mostos de tequila [Mancilla-Margalli, 2002]. La espectrometría de infrarrojo aplicada a muestras de tequila provee por tanto un perfil espectral de la mezcla completa.

Comparados con otros instrumentos tradicionales de infrarrojo un espectrómetro con transformada de Fourier ofrece una mejora significativa de la relación señal/ruido, mayor exactitud de frecuencia, mucha mayor velocidad y procesamiento de datos más eficaz. Una ventaja de la espectrometría con transformada de Fourier es que se puede obtener el interferograma en pocos segundos, si a su vez se realizan decenas o cientos de barridos y se promedian se mejora aún más la relación señal/ruido. Promediar barridos es posible debido a la reproducción precisa de la posición de número de onda entre cada uno de los barridos. Promediando n espectros, la relación señal/ruido mejora \sqrt{n} veces, por tanto entre más espectros se promedien la relación señal/ruido será aún mejor [Harris, 2010].

El espectro neto de transmisión se obtiene al dividir la transformada de la muestra entre la transformada del fondo [Harris, 2010].

La espectrometría de infrarrojo es una técnica muy versátil ya que en muchos casos no es necesario realizar un pretratamiento a las muestras. Permite obtener una gran cantidad de información por cada espectro obtenido. Requiere pocas cantidades de muestra y es una técnica no destructiva.

4.7 Análisis Multivariante de datos o Quimiometría

Los investigadores están acostumbrados a realizar mediciones de sus sistemas de estudio con la finalidad de recabar datos que después puedan analizar e interpretar y a partir de eso obtener conclusiones lógicas. Sin embargo, por mucho tiempo estas mediciones se han realizado de manera individual. Por ejemplo, midiendo números de onda particulares, al igual que su tratamiento posterior.

Con el avance tecnológico de las últimas décadas se ha logrado obtener grandes cantidades de información de un mismo fenómeno. El análisis multivariante tiene por objeto el estudio estadístico de varias variables medidas en los elementos de una población. Sus objetivos principales son:

- Resumir las variables originales en unas pocas nuevas variables con la menor pérdida de información.
- Encontrar grupos de muestras o variables, si es que existen.
- Clasificar muestras nuevas en grupos ya identificados
- Relacionar conjuntos de variables.

Existen métodos multivariantes que están encaminados a extraer información útil de los datos originales cuando de ellos no se conoce nada *a priori*, estos se denominan comúnmente como métodos de exploración de datos o “minería de datos”; esta última denominación indica la capacidad de estas técnicas para extraer información a partir de la materia prima de datos [Peña, 2002]. Ejemplos de estos métodos son el análisis PCA y el análisis Cluster. Métodos como estos que involucran agrupamiento y clasificación son los que constituyen el reconocimiento de pautas [Otto, 2017].

Por otra parte, están los métodos que buscan sacar conclusiones de la población que ha generado los datos, estos requieren la construcción de un modelo que explique su generación y pueda predecir datos futuros. Estos se conoce como métodos de inferencia [Peña, 2002].

Los métodos multivariantes son herramientas muy poderosas para el análisis de datos complejos. Sin embargo es importante entender que muchas veces no basta con realizar un análisis único sino que es necesario conjuntar varias técnicas con la finalidad de tener un panorama mayor del problema estudiado y poder formular conclusiones más acertadas.

El análisis descriptivo se recomienda aplicarlo como primer paso para comprender la estructura de los datos y a partir de los resultados obtenidos pasar a los métodos más complejos.

Durante el análisis descriptivo es posible determinar la homogeneidad de la población o si corresponden a poblaciones distintas que deben estudiarse por separado. Un caso especial de heterogeneidad es el caso de muestras o datos anómalos (outliers). La identificación de estos puntos anómalos es muy importante, ya que sus valores distorsionan los valores descriptivos del conjunto [Peña, 2002]. El término quimiometría fue acuñado en 1972 y esta puede ser definida como la disciplina química que usa métodos matemáticos, estadísticos y de lógica formal para diseñar o seleccionar metodologías de medición y experimentos además de proveer el máximo de información química relevante para el análisis químico [Massart et. al., 2003].

La quimiometría usa una amplia cantidad de métodos multivariantes para lograr la obtención de información a partir de los datos originales. Es importante entender que no es lo mismo tener datos que tener información, los datos pueden ser solo valores numéricos pero estos se convierten en información útil cuando cobran significado real y es posible interpretarlos para obtener conclusiones de ellos.

4.8 Matriz de datos

Para poder comenzar con los métodos multivariantes aplicados en quimiometría es necesario tener información de múltiples variables de diferentes muestras analizadas. Esta información se recoge en una tabla donde se colocan de manera horizontal las variables y de forma vertical cada una de las muestras analizadas. Las hojas de cálculo como Excel® permiten la creación de matrices de gran tamaño.

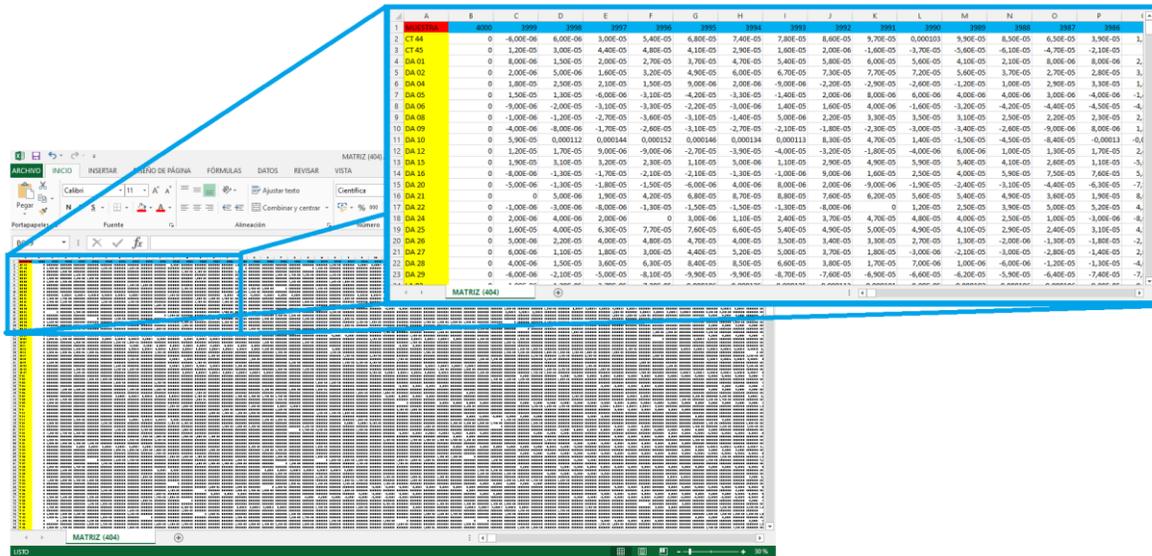


Figura 13.- Representación de una Matriz de datos en una hoja de cálculo de Excel [Excel®, 2013].

Si se miden p variables numéricas en un conjunto de n elementos. Cada una de estas p variables se denomina una variable vectorial o univariable y el conjunto de las p variables forman una matriz o conjunto multivariable. Los valores de las p variables vectoriales en cada uno de los n elementos (lo que se conoce como dato escalar o, simplemente, dato) pueden representarse en una matriz, \mathbf{X} , de dimensiones $(n \times p)$, que llamaremos matriz de datos [Peña, 2002].

4.9 Pretratamiento de datos

En quimiometría es común escuchar la frase “mete basura y obtendrás basura”; a pesar de lo directa que pueda ser esta frase no deja de ser completamente verdadera. No basta con tener cientos de datos si estas mediciones no se analizan correctamente. El primer paso suele consistir en eliminar información no esencial y/o corregir efectos que pueden venir generados en la medida (p.ej., líneas de base no constantes). Para ello se realiza un pretratamiento de los datos.

Un pretratamiento es una serie de pasos de cálculo encaminados a minimizar las fluctuaciones (fuentes de variación o varianza) de las mediciones no inherentes a las muestras.

Existen muchas formas diferentes de pretratamiento. En FT-MIR este pretratamiento consiste en las conversiones de absorbancia a transmitancia, correcciones ATR, corrección de línea base (automática o interactiva), suavizados, derivadas, entre otros.

4.10 Escalados.

Existen diferentes métodos de escalado y los softwares actuales permiten la selección de diferentes de ellos de manera automática. Estos algoritmos se aplican una vez que se trataron los datos brutos con los pretratamientos adecuados. Se trata de disponer los datos en la mejor situación de partida para su análisis posterior. Dos de los más comunes son el Centrado en la media y el Autoescalado.

4.10.1 Centrado en la media.

Para eliminar un desplazamiento constante, los datos se pueden trasladar a lo largo del origen de coordenadas. El procedimiento común es mediante el centrado donde cada variable se centra al sustraerle el valor de la media de la columna correspondiente [Otto, 2017].

$$X_{ik}^* = X_{ik} - \bar{X}_k \quad \text{donde:}$$

i = referencia a la fila

k = referencia a la columna

\bar{X}_k = media de la columna

Donde la media de la columna está calculada por:

$$\bar{X}_k = \frac{1}{n} * \sum_{i=1}^n X_{ik} \quad \text{donde:}$$

n = numero de muestras

4.10.2 Autoescalado

El autoescalado se define matemáticamente por las siguientes ecuaciones [Otto, 2017].

$$X_{ik}^* = \frac{X_{ik} - \bar{X}_k}{S_k} \quad \text{donde:}$$
$$S_k = \sqrt{\frac{\sum_{i=1}^n (X_{ik} - \bar{X}_k)^2}{n - 1}}$$

El autoescalado da como resultado datos con media cero y varianza igual a la unidad. La longitud de los vectores se escala a $\sqrt{n - 1}$.

La figura 14 muestra los efectos de los diferentes escalados en un conjunto de datos.

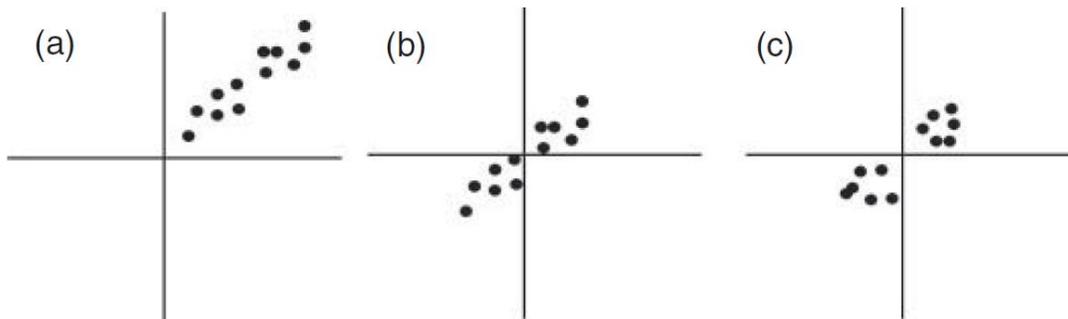


Figura 14.- a) Sin Escalado, b) Centrado en la media, c) Autoescalado [Otto, 2017 a].

4.11 Técnicas de agrupación no supervisadas

El agrupamiento de datos analíticos (muestras o variables) es posible por medio de los métodos de agrupamiento, como tales, o por medio de proyecciones de los datos en espacios dimensionales inferiores. Este tipo de representaciones son de gran utilidad ya que permiten visualizar el comportamiento de las muestras (variables) que de otra manera sería muy difícil de identificar. Estos se conocen como métodos no supervisados debido a que se desconoce la pertenencia de las muestras a los diferentes grupos existentes [Otto, 2017].

4.11.1 Análisis de Componentes Principales (PCA).

El análisis de componentes principales (PCA) es con seguridad la técnica multivariante más usada y de mayor antigüedad. Esta técnica es atribuida a Hotelling (1933) aunque su origen se encuentra en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson (1901) [Peña, 2002].

La idea básica del método consiste en describir la variación de un conjunto de datos multivariantes en términos de un conjunto de variables no correlacionadas donde cada una de ellas es una combinación lineal de las variables originales [Everitte, 2001]. Estas combinaciones son los denominados componentes (o factores) principales a los cuales también se les conoce, menos frecuentemente, como variables latentes.

Matemáticamente el análisis PCA representa a la matriz \mathbf{X} como un producto de dos matrices más pequeñas, la matriz de Scores y la matriz de Loadings. La figura 15 muestra una representación de esta idea.

$$X = T L^T$$

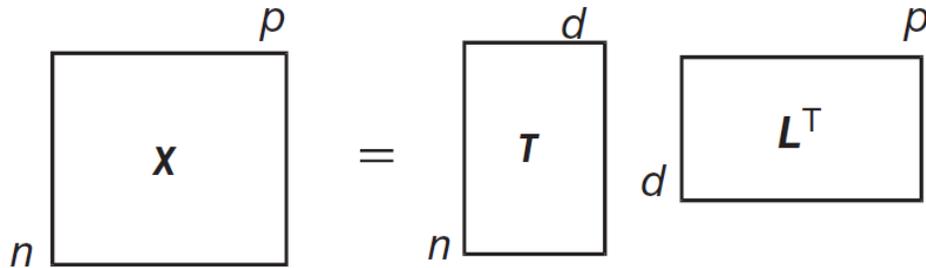


Figura 15.- Representación gráfica de la matriz X en PCA [Otto, 2017 b].

Donde X es la matriz original con n filas (muestras) y p columnas (variables), T es la matriz de scores con n filas y d columnas (componentes principales) y L es la matriz de loadings con d columnas y p filas; T es la transpuesta de la matriz L .

Los componentes principales (PCs) pueden ser considerados como proyecciones de la matriz original, por lo tanto la información mostrada en la figura 15 puede representarse como se muestra en la figura 16.

$$T = X L$$

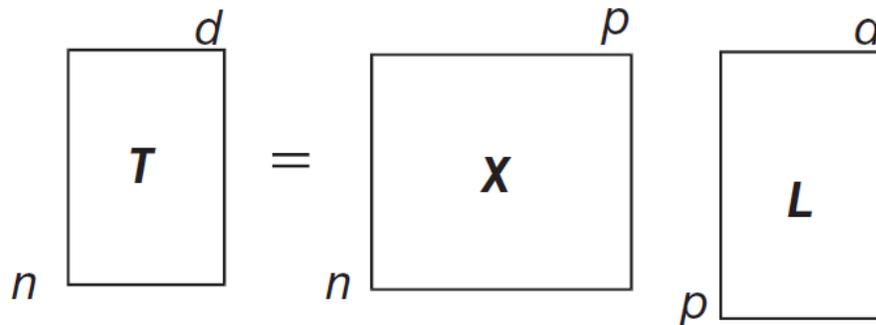


Figura 16.- Representación gráfica de los PCs [Otto, 2017 c].

Muchos de los análisis multivariantes buscan en primer lugar una reducción de la dimensionalidad de los datos cuando está presente la correlación. Esto quiere decir que se logre describir con precisión la disposición original de las muestras por medio de un pequeño subconjunto de ellas (o por medio de un número reducido de factores que se calculan a partir de las variables originales).

En el caso de PCA la reducción de dimensionalidad se logra por medio de la construcción de los componentes principales.

Otra forma de explicar los componentes principales (Z_1, Z_2, \dots, Z_n) es considerar que se obtienen por medio de combinaciones lineales de las variables originales (X_1, X_2, \dots, X_n) de la siguiente forma:

$$Z_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \dots + a_{1n}X_n$$

$$Z_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \dots + a_{2n}X_n$$

Etc.

Los coeficientes “ a_{nn} ” se eligen de manera que las nuevas variables (PCs) no estén correlacionadas [Miller, 2002].

Los PCs son determinados en base al criterio de máxima variación. Esto quiere decir que la mayor varianza de los datos está contenida en el primer componente principal. Cada PC subsecuente describe el máximo de varianza que aún no es explicada por los PCs formados previamente [Otto, 2017]. Esto lleva a que cuando haya correlación significativa en las variables originales el número de PCs útiles será mucho menor que el número de variables originales.

Una propiedad interesante de los PCs es que forman ángulos rectos entre sí, esta propiedad es conocida como ortogonalidad [Miller, 2002].

Muchas veces el primer componente principal puede satisfacer los requerimientos descriptivos del investigador, sin embargo no siempre este componentes es el más interesante [Everitte, 2001].

Usar todos los componentes principales no está justificado. Esto debido a que por lo general es necesario separar los PCs “puros” (más relevantes) de aquellos que introducen ruido. Para decidir cuantos PCs utilizar existen diversos métodos aproximados, entre los más comunes se encuentran los siguientes:

1. Porcentaje de varianza explicada.
2. Criterio del valor propio (o eigenvalor) unidad.

El método de porcentaje de varianza explicada se aplica cuando el experimentador ha trabajado con conjuntos de datos similares y ha ganado suficiente experiencia con ello. Usualmente un 90 % de varianza explicada es suficiente.

El criterio de valor propio uno se basa en la idea de que el valor propio (eigenvalor) promedio para datos autoescalados es uno. Esto resulta del hecho de que en datos estandarizados la suma de los valores propios de la matriz de correlación es exactamente el número de variables. Por lo tanto solo PCs con valores propios mayores a la unidad son considerados.

4.11.2 Análisis Cluster.

El análisis Cluster o de agrupaciones es una técnica de clasificación (estrictamente, de agrupación) no supervisadas ya que, como en PCA, no puede hacerse uso del posible conocimiento que se tenga de antemano de la pertenencia de las muestras a los grupos. Es una técnica que puede ser muy útil cuando PCA no da buenos resultados.

Las muestras se agregan a un grupo (cluster) de acuerdo a su similitud con otras muestras en función de sus características. Esto da como resultado la formación de clusters (grupos), que se pueden obtener usando métodos jerárquicos y no jerárquicos [Otto, 2017].

Un cluster describe un grupo de objetos que son mucho más similares entre ellos que con otros objetos que se encuentran fuera del cluster. Su formación comienza con un objeto semilla a partir del cual se agregan los demás objetos [Otto, 2017].

La similitud entre muestras se evalúa por medio de la distancia entre ellas. La idea es simple, entre menor sea la distancia entre dos muestras mayor será la similitud que compartan. Existen diversos métodos de calcular estas distancias, la tabla 1 recopila algunos de ellos.

Tabla 1. Distancias comunes en análisis Cluster y sus ecuaciones.

Distancia	Ecuación
Euclidea	$d = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_n - Y_n)^2}$
Euclidea cuadrada	$d = (X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_n - Y_n)^2$
Manhattan	$d_{ij} = \sum_{k=1}^K X_{ik} - X_{jk} $
Pearson	$d_{ij} = \sqrt{\frac{\sum_{k=1}^K (X_{ik} - X_{jk})^2}{S_j^2}}$
Mahalanobis	$D_{ij}^2 = (X_i - X_j)^t C^2 (X_i - X_j)$

Generalmente los softwares tienen la posibilidad de elegir entre varias opciones para calcular la distancia entre muestras. En este trabajo únicamente se mencionan las ecuaciones para dar una panorama de cómo funcionan estos cálculos, sería complicado tratar de explicar detalladamente cada una de ellas además de poco práctico ya que actualmente no se realizan de forma manual.

Calcular las distancias es un paso muy importante, sin embargo no es suficiente para conducir a agrupamientos. El siguiente paso es elegir un método de agrupación, también llamado algoritmo cluster. Haciendo uso de las distancias calculadas anteriormente y de dicho algoritmo se pueden agrupar las muestras en sus grupos correspondientes. Existen muchos algoritmos cluster, entre los principales se encuentran los siguientes:

1. *Weighted Average Linkage* (Enlace promedio ponderado): En este método se asume que los pesos y tamaños de los cluster son iguales. La distancia entre clusters se calcula de la siguiente manera:

$$d_{ki} = \frac{d_{Ai} + d_{Bi}}{2}$$

2. *Single Linkage* (Enlace único): en este método la distancia entre clusters opuestos se calcula por medio de la siguiente ecuación:

$$d_{ki} = \frac{d_{Ai} + d_{Bi}}{2} - \frac{|d_{Ai} - d_{Bi}|}{2} = \min(d_{Ai}, d_{Bi})$$

Gráficamente la ecuación anterior significa que la unión entre clusters se realiza tomando en cuenta la mínima distancia posible entre los clusters tomando como punto de referencia la muestra más cercana al cluster opuesto, la figura 17 muestra cómo se realiza gráficamente.

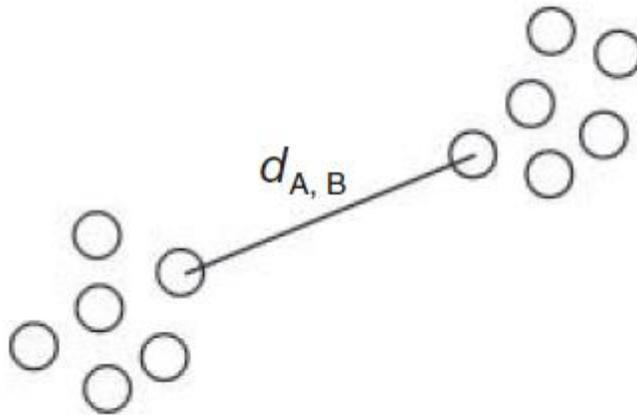


Figura 17.-Unión de Clusters con distancia Single Linkage.

Este método tiende a formar clusters alargados y con cierto grado de heterogeneidad.

3. *Complete Linkage* (Enlace completo): este método puede considerarse como un opuesto al método de Single Linkage ya que las uniones entre clusters las logra considerando la máxima distancia entre objetos:

$$d_{ki} = \frac{d_{Ai} + d_{Bi}}{2} + \frac{|d_{Ai} - d_{Bi}|}{2} = \max(d_{Ai}, d_{Bi})$$

Tiende a formar clusters esféricos.

4. *Ward's Algorithm* (Algoritmo de Ward): en este método los clusters son agregados de manera que se minimice el incremento de error en la suma de cuadrados:

$$d_{ki} = \frac{n_A + n_i}{n + n_i} d_{Ai} + \frac{n_B + n_i}{n + n_i} d_{Bi} - \frac{n_i}{n + n_i} d_{AB}$$

Este método logra impedir que crezca la varianza interna de los clusters y que la varianza externa (varianza entre clusters) aumente lo que da como resultado que se diferencien mejor y se obtengan resultados más fácilmente interpretables. Genera gráficos bien estructurados por lo que es el método más utilizado.

La mejor forma de ver los resultados de un análisis cluster es por medio de un dendrograma el cual es un gráfico que permite visualizar en uno de sus ejes las muestras analizadas y en el otro las similitudes (distancias) entre ellas. Es común que se refiera al dendrograma también como árbol de clasificación, la figura 18 da una imagen típica de un dendrograma.

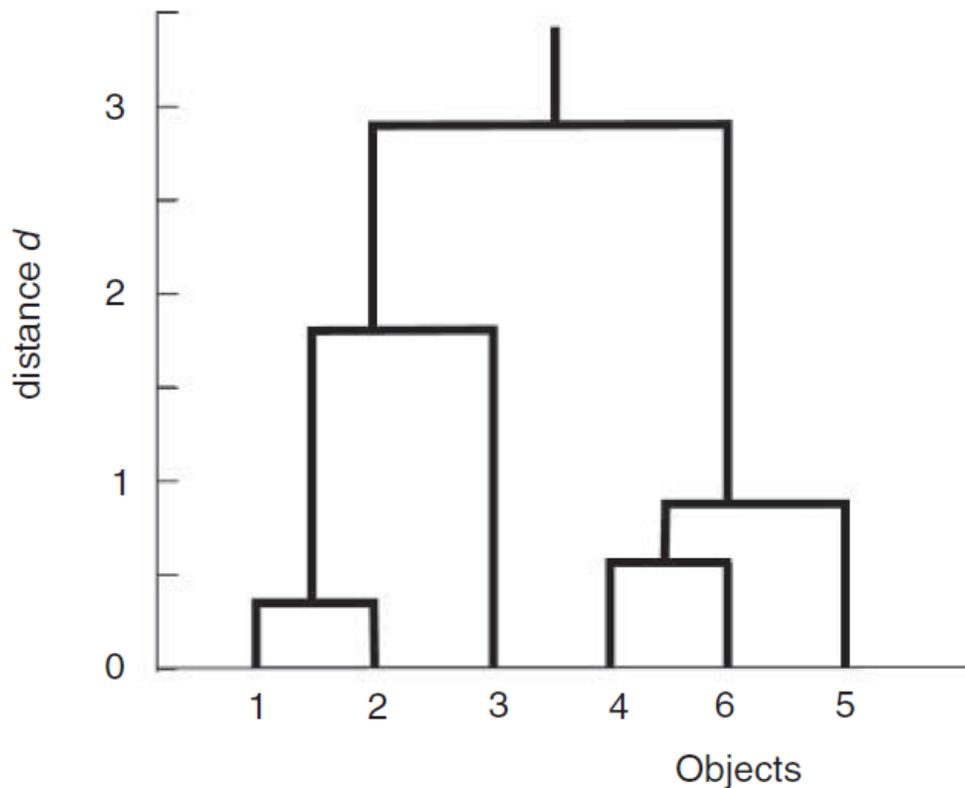


Figura 18.- Forma típica de un dendrograma.

Un dendrograma muestra las uniones entre muestras y después entre cada uno de los cluster formados hasta terminar con un único cluster conteniendo a todas las muestras analizadas, sin embargo llegar hasta el punto de tener un único cluster carece de sentido. El número de clusters correcto dependerá de lo que más convenga al experimentador y así mismo el será el responsable de decidir dónde debe ser cortado el árbol de clasificación.

4.12 Precauciones con las técnicas de agrupamiento no supervisadas.

Es importante recordar que las técnicas de agrupamiento no supervisadas tienen por principal objetivo obtener un panorama general de las muestras analizadas. Al trabajar con estas técnicas lo mejor es ir probando las diferentes combinaciones posibles; escalados, distancias, métodos, etc. hasta encontrar el modelo que sea más satisfactorio. Cambiar cualquier parámetro en los análisis dará resultados diferentes lo cual se verá claramente reflejado en los gráficos obtenidos; sin embargo, no por que éstos sean diferentes significa que sean resultados erróneos. La diferencia entre un resultado y otro será la facilidad de interpretación que se obtenga de éste.

4.13 Técnicas de clasificación supervisadas

Después de realizar un primer análisis valiéndose de las técnicas exploratorias es posible utilizar unos métodos un poco más complejos como lo son las técnicas de clasificación supervisadas. A diferencia de las técnicas no supervisadas en este caso se sabe anticipadamente la pertenencia de las muestras a los diferentes grupos. Estas técnicas están encaminadas a separar los diferentes grupos mediante la búsqueda de una regla de clasificación que permita ubicar las diferentes muestras en su grupo correcto.

Encontrando la regla de clasificación correcta es posible dar un carácter predictivo a estas técnicas ya que permiten predecir la pertenencia de una muestra desconocida a un grupo particular.

Existen dos categorías de muestras en el análisis supervisado. La primera de ellas se constituye por las muestras de entrenamiento, también conocidas como muestras de aprendizaje. La segunda viene dada por las muestras de prueba. Con las muestras de entrenamiento se consigue la regla de clasificación y el modelo obtenido es el que permite realizar la posterior predicción de las muestras de prueba.

El siguiente paso consiste en comprobar que el modelo obtenido es “bueno” y que proporciona resultados satisfactorios y verdaderos. Para esto se debe evaluar el modelo poniendo a prueba su capacidad de reconocimiento (clasificación) y de predicción.

La capacidad de reconocimiento se evalúa de acuerdo al porcentaje de muestras de entrenamiento que son correctamente asignadas a su grupo. Por otra parte la capacidad de predicción se evalúa de acuerdo al porcentaje de muestras de prueba que son asignadas correctamente a su grupo. Sin embargo aquí surge un problema, una muestra de prueba en teoría sería aquella para la cual el modelo ignora su identidad y, por tanto, su pertenencia a algún grupo en específico; evaluar la predictibilidad del modelo se vuelve incierto si se realiza de esta manera [Massart, 2003].

Uno de los métodos más habituales para evaluar la predictibilidad de un modelo es el método de validación por cruces o validación cruzada. Este método consiste en retirar

algunas de las muestras de entrenamiento y asignarlas como muestras de prueba, aquí se tiene la ventaja de que se conoce la identidad de las muestras de prueba. El procedimiento se realiza variando las muestras de prueba hasta que todas las muestras de entrenamiento hayan pasado una vez como muestra de prueba. Con toda la información obtenida se calcula el porcentaje de muestras de prueba que fueron clasificadas correctamente y de esta manera se evalúa la predictibilidad del modelo [Massart, 2003].

Un error que se suele cometer es que después de obtener un buen resultado de capacidad de reconocimiento se omite la evaluación de la capacidad de predicción. Es importante destacar que las capacidades de reconocimiento y predicción dan información importante de la eficiencia del modelo y que estas están estrechamente relacionadas entre sí. Sin embargo un modelo puede realizar un correcto reconocimiento pero una mala predicción.

Si después de evaluar el modelo los resultados de evaluaciones de reconocimiento y predicción son sustancialmente distintos significara que la regla de clasificación obtenida depende demasiado del conjunto de muestras de entrenamiento usadas; los resultados obtenidos de esta manera no son estables y no pueden ser confiables.

4.13.1 Curvas de Potencia

En esta técnica, una vez generado el modelo con las muestras de entrenamiento, se asigna cada una de las muestras de prueba al grupo con el que comparte mayor similitud, según el cálculo de una probabilidad de pertenencia a cada uno de los grupos predefinidos.

Esta técnica se encuentra dentro del grupo de técnicas conocidas como “funciones de densidad” o “funciones de probabilidad”. Todas ellas tienen la característica de limitar una región espacial para cada grupo. Al proyectar las muestras de prueba sobre el plano de trabajo (en esta versión del programa Datan, PC1-PC2) se calculan las probabilidades de pertenencia a cada uno de los grupos existentes y se asignan al grupo donde presenten los mayores porcentajes de similitud.

En las técnicas de funciones de densidad se considera un campo de probabilidad (o potencial) alrededor de las muestras de entrenamiento que formen un grupo. Esta es la razón de por qué se conocen como métodos de potencia.

Siempre es recomendable el uso de funciones simples, como funciones triangulares o gaussianas. La función seleccionada está caracterizada por su amplitud.

Para las funciones potenciales planteadas a partir de puntos individuales, la función potencial acumulativa se determina por adición de las señales de las funciones potenciales individuales en cada posición a lo largo del eje x. Dividiendo la función potencial acumulativa de una clase entre el número de muestras que contribuyeron a su formación

se obtiene la función potencial de la clase. De esta manera se le da un carácter probabilístico el cual permite lograr las clasificaciones probabilísticas [Massart, 2003].

El método de curvas de potencia emplea funciones gaussianas en dos dimensiones. La función de probabilidad con distribución normal está dada por la siguiente ecuación [Massart, 2003].

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \right\} \right]$$

Donde μ_j y σ_j^2 son la media y la varianza de x_j ($j = 1, 2$) y ρ es el coeficiente de correlación.

4.13.2 Clasificación mediante vectores soporte, SVM (*Support Vector Machines*)

La separación de grupos traslapados es muy complicada con métodos habituales, no sólo mediante curvas de potencia sino otros más avanzados, como el análisis discriminante, ya que en estos se busca una separación óptima mediante hiperplanos (es decir, en una dimensionalidad inferior a la de los datos de partida). SVM, por el contrario, provee una alternativa eficiente; una transformación del espacio de las variables que permite construir una frontera en un espacio de dimensiones mayores para separar grupos con fronteras no lineales.

SVM es un caso especial de los métodos basados en Kernel. Su origen se remonta a los trabajos de Vapnik en la década de 1970, esta técnica se basa en dos ideas principales. La primera de ellas consiste en mapear los vectores de características de una forma no lineal hacia un espacio dimensional superior (posiblemente infinito) y entonces usar un clasificador lineal en este nuevo espacio. La segunda idea es la de usar un clasificador lineal de “margen grande” que separe los datos tanto como sea posible entre los hiperplanos (fronteras) [Kulkarni, 2011].

El nombre vectores de soporte viene del hecho que las muestras que se encuentran en la orilla de los márgenes reciben el nombre de vectores soporte ya que son los que se utilizan para calcular dichas fronteras (bordes). Por extensión, el clasificador que se genera con base a estos vectores de soporte se denomina Clasificador de Vectores de Soporte (SVM).

5.0 Desarrollo Experimental.

5.1 Muestras Analizadas.

Las muestras fueron otorgadas por el CRT al laboratorio 10 de la Unidad de Investigación Multidisciplinaria (UIM) de la FES-Cuautitlán. Estas muestras fueron recibidas en botellas de PET con etiquetas que permiten identificarlas. La figura 19 muestra una imagen del aspecto de las muestras.

Para este trabajo se usaron 131 muestras que involucran destilados/licores de agave y tequilas 100 % agave y mixtos en sus clases blanco, joven, reposado, añejo y extra añejo. El anexo 1 ofrece una lista detallada de cada muestra así como los códigos individuales asignados a cada una de ellas.



Figura 19.- Presentación de las muestras utilizadas en este trabajo.

A cada muestra se le asignó una nomenclatura específica que permite identificarla fácilmente. La primera parte del código de identificación corresponde al tipo de muestra, T: tequila, DA: destilado de agave y LA: licor de agave. La segunda parte aplica únicamente para las muestras de tequila, corresponde a las clases de tequila las cuales quedan identificadas de la siguiente manera: B: blanco, J: joven, R: reposado, A: añejo, EA: extra

añejo; adicional a esto si el código presenta a continuación una letra M significa que la muestra es un tequila mixto, en caso contrario se trata de un tequila 100 % agave. La parte final corresponde a un número, el mismo número que otorga el CRT a las muestras.

Para los tratamientos quimiométricos en el software Datan se les asigno un color diferente a cada una de las muestras a menos que se indique lo contrario en algún caso en específico.

Destilados de agave y Licores de agave: Amarillo

Tequilas Blancos: Verde

Tequilas Jóvenes: Azul claro

Tequilas Reposados: Azul oscuro

Tequilas Añejos y Extra añejos: Rojo

5.2 Uso del Espectrómetro FT-IR

El equipo utilizado fue un espectrómetro de la marca Perkin Elmer® modelo FT-IR Frontier. Las mediciones de los espectros se obtuvieron con el uso del instrumento de Reflexión Total Atenuada (ATR) de la marca Perkin Elmer®.

1. Ajustar las condiciones del equipo bajo las siguientes especificaciones:
 - a) Intervalo de $4000\text{ cm}^{-1} - 400\text{ cm}^{-1}$.
 - b) 16 barridos.
 - c) Resolución 4 cm^{-1} .
 - d) Fondo entre cada muestra.
2. Para limpiar, tomar aproximadamente $50\ \mu\text{L}$ de etanol con una micro pipeta y limpiar con ayuda de papel seda el diamante del ATR. Repetir 2 veces.
3. Realizar un barrido de fondo usando fondo aire.
4. Tomar una pequeña porción de muestra en un vaso de precipitados limpio.
5. Lavar la punta de la micro pipeta 3 veces con la muestra que se medirá desechando las porciones usadas en el lavado.
6. Tomar $50\ \mu\text{L}$ de muestra y limpiar el ATR con ayuda de papel seda.
7. Tomar otros $50\ \mu\text{L}$ de muestra y colocarlos sobre el diamante del ATR con el cuidado de cubrirlo completamente y evitando burbujas en el interior de la gota.
8. Realizar el barrido de la muestra para obtener el espectro en transmitancia.
9. Repetir los pasos 2-8 con cada una de las muestras.
10. Guardar los archivos de los espectros en formato .ASCII y .sp.

Nota: Para reducir el error sistemático de las mediciones, las muestras se midieron aleatoriamente, por el mismo analista.

Las mediciones de las muestras se realizaron en un intervalo de tiempo de una semana con la finalidad de reducir el error de derivas causadas por realizar mediciones en fechas muy distantes unas de otras. También se realizaron en un ambiente de temperatura y humedad controlada.

5.3 Pretratamiento de los datos.

Los espectros obtenidos en transmitancia deben ser adecuados para el análisis quimiométrico, para esto es necesario realizar un pretratamiento. Este pretratamiento consiste en convertir los datos de transmitancia a valores de absorbancia. Realizarles una corrección ATR, una corrección de línea base (CLB) automática o interactiva (según elección) y el ajuste a cero.

La figura 20 muestra el proceso del pretratamiento de datos así como su efecto sobre los espectros.



Figura 20.- Efecto de los diferentes pretratamientos en los datos FIT-MIR.

La corrección ATR permite corregir algunos efectos físicos debidos a la reflectancia total, de manera que los espectros transformados sean mucho más parecidos a los espectros obtenidos por transmisión. La diferencia más significativa entre un espectro de reflexión interna y uno de transmisión se relaciona con las intensidades relativas de las bandas. Esto debido a que un espectro obtenido con ATR muestra intensidades mejoradas de las bandas en longitudes de onda más largas [Spectrum, 2015].

La corrección de línea base permite eliminar las pendientes del espectro causadas principalmente por dispersión del haz de infrarrojo de la muestra y los efectos de fondo. Esto permite una mejor comparación entre espectros [Spectrum, 2015].

Todos estos pretratamientos se realizaron con el software Spectrum® el cual está incluido con el equipo de infrarrojo marca Perkin Elmer®.

5.4 Construcción de matrices.

Las matrices se construyeron en Excel® [Microsoft Excel, 2013] a partir de los espectros guardados en formato ASCII. Se copian los datos de cada uno de los espectros en un archivo nuevo de Excel® [Microsoft Excel, 2013] de modo que al finalizar presente el formato como el mostrado en la figura 21.

Muestra	4000	3999	3998	3997	3996	3995	3994	3993	3992	3991	
DA 02	0,000772	0,000776	0,00078	0,000793	0,000811	0,000829	0,000842	0,000851	0,000859	0,000864	0,000868
DA 04	0,000236	0,000256	0,000265	0,000263	0,000258	0,000255	0,00025	0,000241	0,00023	0,000224	0,00022
DA 05	0,000573	0,000591	0,000592	0,000575	0,000553	0,000545	0,000556	0,000578	0,000596	0,000604	0,000608
DA 06	0,000532	0,000526	0,000518	0,000511	0,000511	0,000524	0,000547	0,000566	0,000571	0,000561	0,000556
DA 08	0,00135	0,001352	0,001342	0,001329	0,001322	0,001328	0,001346	0,001366	0,001384	0,001395	0,001406
DA 09	-0,000184	-0,000184	-0,000185	-0,000191	-0,000197	-0,000199	-0,000193	-0,000184	-0,000179	-0,000181	-0,000186
DA 10	0,000817	0,000886	0,000951	0,000993	0,001012	0,001018	0,001016	0,001006	0,000987	0,000962	0,000937
DA 12	0,000071	0,000087	0,000095	0,00009	0,000076	0,000061	0,000052	0,000054	0,000065	0,000081	0,000096
DA 15	0,001045	0,001065	0,001079	0,001082	0,001074	0,001063	0,001059	0,001067	0,001086	0,001107	0,001128
DA 16	0,000875	0,000869	0,000868	0,000865	0,000863	0,000866	0,000876	0,000889	0,000902	0,000912	0,000922
DA 20	0,001189	0,001186	0,001181	0,001179	0,001184	0,001197	0,00121	0,001216	0,001213	0,001204	0,001195
DA 21	0,000988	0,000989	0,000996	0,001011	0,001035	0,001063	0,001083	0,001086	0,001075	0,001062	0,001049
DA 22	0,000978	0,000979	0,000979	0,000976	0,000973	0,000973	0,000975	0,000979	0,000986	0,000996	0,001006
DA 24	0,000709	0,000713	0,000717	0,000718	0,000719	0,000724	0,000734	0,000749	0,000766	0,000778	0,00079
DA 25	0,001501	0,001518	0,001543	0,001568	0,001583	0,001583	0,001574	0,001564	0,001559	0,001561	0,001562
DA 26	0,001511	0,001518	0,001538	0,001557	0,001568	0,001568	0,001564	0,001564	0,001562	0,001562	0,001562

Figura 21.- Formato de una matriz en Excel®.

5.5 Datan®

Finalmente, la matriz generada en Excel® [Microsoft Excel, 2013] se copió al software Datan® [Datan, 2013] donde se procedió a realizar los análisis quimiométricos. Aquí fue necesario guardar los archivos en los formatos .mdf y .dpx. La figura 22 muestra la vista general de Datan® [Datan, 2013] con una matriz de datos cargada.

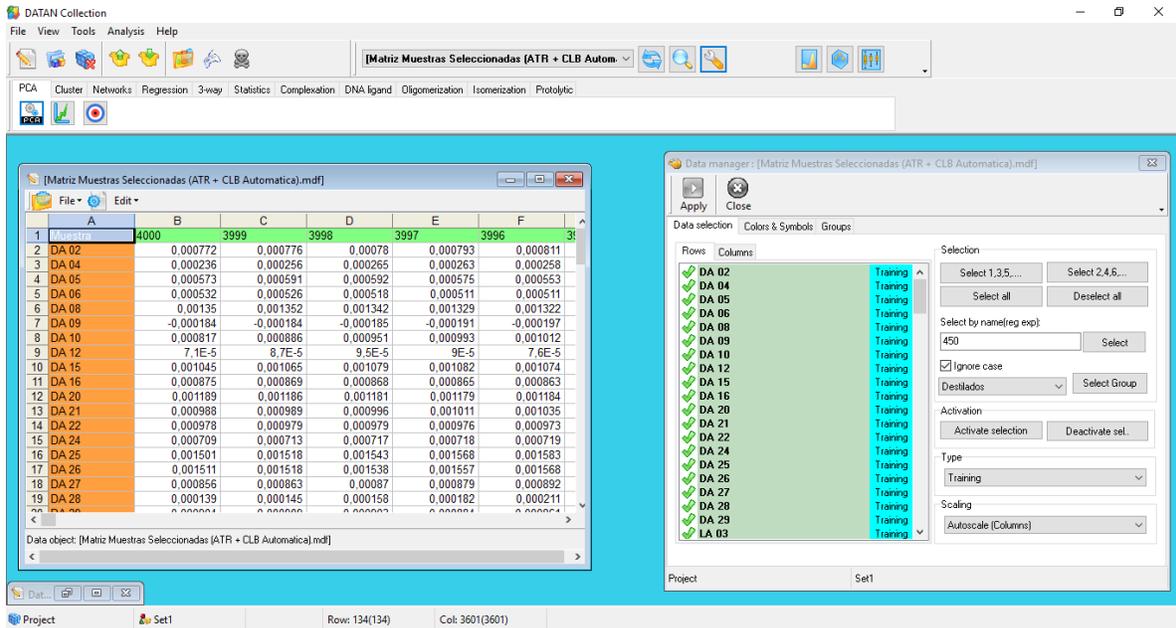


Figura 22.- Vista previa de Datan® con una matriz de datos.

En este software se hizo uso de los análisis PCA, Cluster, SVM y Curvas de Potencia. Se eligió este software debido a que a pesar de no ser muy conocido fue el que permitió trabajar con la extensa cantidad de datos contenidos en las matrices.

6.0 RESULTADOS OBTENIDOS AL TRATAR ESPECTROS ATR CON CORRECCIÓN DE LÍNEA BASE AUTOMÁTICA

En el presente estudio se construyeron modelos para diferenciar tequilas auténticos de los destilados de agave a partir de datos FT-MIR por medio de diversas técnicas quimiométricas supervisadas (Curvas de potencia y Support Vector Machines) y no supervisadas (Análisis de Componentes Principales y Cluster).

Muestras: La matriz definitiva se conformó por 131 muestras totales incluyendo 10 licores de agave, 18 destilados de agave, 25 tequilas blancos, 15 tequilas jóvenes, 24 tequilas reposados, 25 tequilas añejos y 14 tequilas extra añejos. Los licores y destilados se consideran en un sólo grupo, al igual que los añejos y extra añejos en otro. Todas las muestras fueron suministradas por el CRT.

Las muestras para el proceso de aprendizaje se seleccionaron a partir de un PCA previo para diferenciación de clases de forma que las muestras fueran representativas de su clase. Esta matriz definitiva se denominó “matriz muestras seleccionadas”.

Vale la pena mencionar que en todos los casos donde se intentó construir modelos para diferenciar (o “discriminar”, usando este término en sentido laxo) entre tequilas y licores/destilados, ya sea por técnicas no supervisadas como supervisadas, el número de muestras de cada conjunto está desbalanceado, es decir, el número de muestras de licores/destilados es de 28, en tanto que el de tequilas es de 103. No obstante, se procedió a intentar construir los modelos.

Los espectros fueron obtenidos como se indicó anteriormente. Se convirtieron de Transmitancia a Absorbancia y se realizó la corrección ATR, seguida de Corrección de Línea Base Automática (CLB) con ajuste a cero. Este punto fue ampliamente explorado probando diferentes opciones, con CLB interactiva, y usando diversos conjuntos de puntos para la corrección. Sin embargo, se observó que los PCA era muy susceptibles a este procedimiento, por lo que se consideró que los modelos resultantes no eran robustos. En el caso de decidirse por esta modalidad, nuestra experiencia invita a usar un número mínimo de puntos de corrección CLB, incluyendo los números de onda inicial y el final para tratar de conservar el aspecto de los espectros originales. Por lo anterior, se decidió usar la corrección de línea base automática con ajuste a cero para evitar sesgar los modelos.

6.1 Técnicas no supervisadas o de exploración de pautas.

6.1.1 Análisis de Componentes Principales (PCA)

6.1.1.1 Modelos para la diferenciación de tequilas vs licores/destilados

Primeramente, se realizó un PCA con el intervalo completo ($4000\text{-}400\text{ cm}^{-1}$) sin ningún escalado en los datos (*none*). Para facilitar la visualización de muestras se presentan en círculo naranja los destilados/licores de agave, y con un cuadrado púrpura a los tequilas de todas las clases.

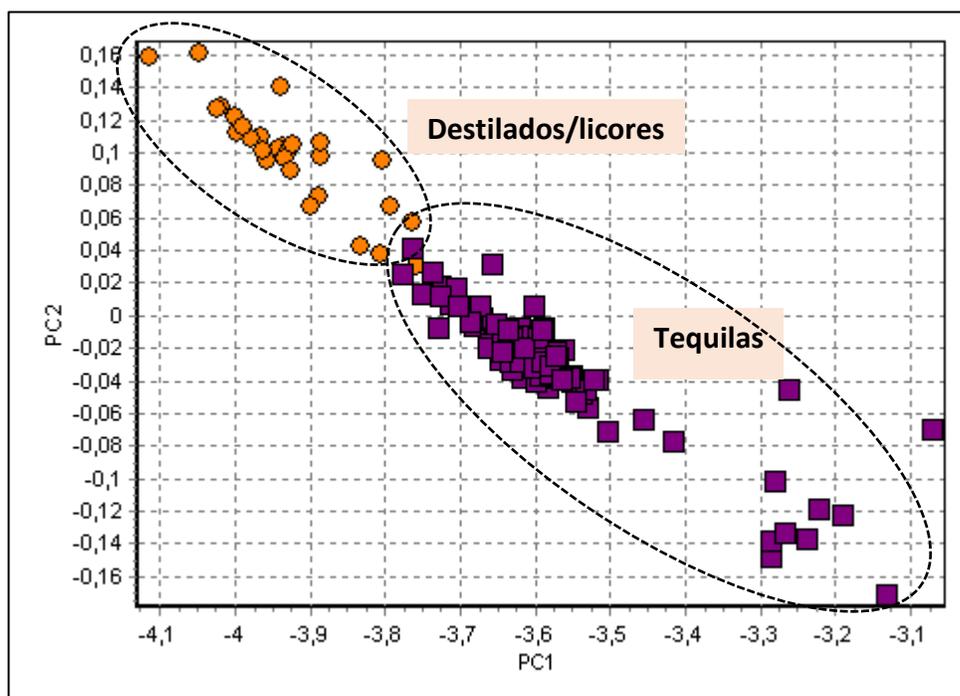


Figura 23.- Gráfico de scores para el conjunto de licores/destilados y tequilas en el intervalo: $400\text{-}4000\text{ cm}^{-1}$, Sin escalado. PC1: 99.9571 %, PC2: 0.0306 %, PC3: 0.0079 %.

De la figura 23 se puede observar que sin ningún escalado y considerando el intervalo espectral completo existe una clara tendencia a la separación de ambos grupos (tequilas y destilados). Se aprecia de forma descendente en diagonal de izquierda a derecha los scores conformados por destilados/licores de agave, seguidos de los tequilas de todas las clases y categorías.

6.1.1.2 Selección de intervalo de número de onda

Después de estudiar el primer PCA obtenido sin escalado y con el intervalo espectral completo se probaron diferentes intervalos con diferentes escalados para determinar el óptimo del modelo. A partir de la observación del espectro se combinaron dos intervalos,

de $980\text{-}1800\text{ cm}^{-1}$ y $2800\text{-}3200\text{ cm}^{-1}$. Posteriormente se decidió realizar un análisis preliminar de loadings para confirmar estos intervalos. Este análisis mostró la existencia de señales en los intervalos de $800\text{ - }1800\text{ cm}^{-1}$ y entre $2800\text{ - }3600\text{ cm}^{-1}$; dichos intervalos abarcan el intervalo óptimo ($980\text{-}1800\text{ cm}^{-1}$ y $2800\text{-}3200\text{ cm}^{-1}$) usado en los análisis posteriores (ver figura 24).



Figura 24.- Loadings asociados a los primeros componentes principales en el intervalo de $400\text{ - }4000\text{ cm}^{-1}$.

La figura 25 muestra los perfiles de los espectros de las muestras analizadas con diferentes escalados, es decir, sin escalado, centrado en la media y autoescalado. Se han resaltado en color rosa los intervalos óptimos de trabajo usados en los análisis posteriores. También es importante observar el efecto de los diferentes escalados; visualmente se aprecia como con los escalados centrado en la media y autoescalado se hacen patentes ciertas diferencias entre las muestras. Los posteriores análisis ponen en evidencia esta primera impresión.

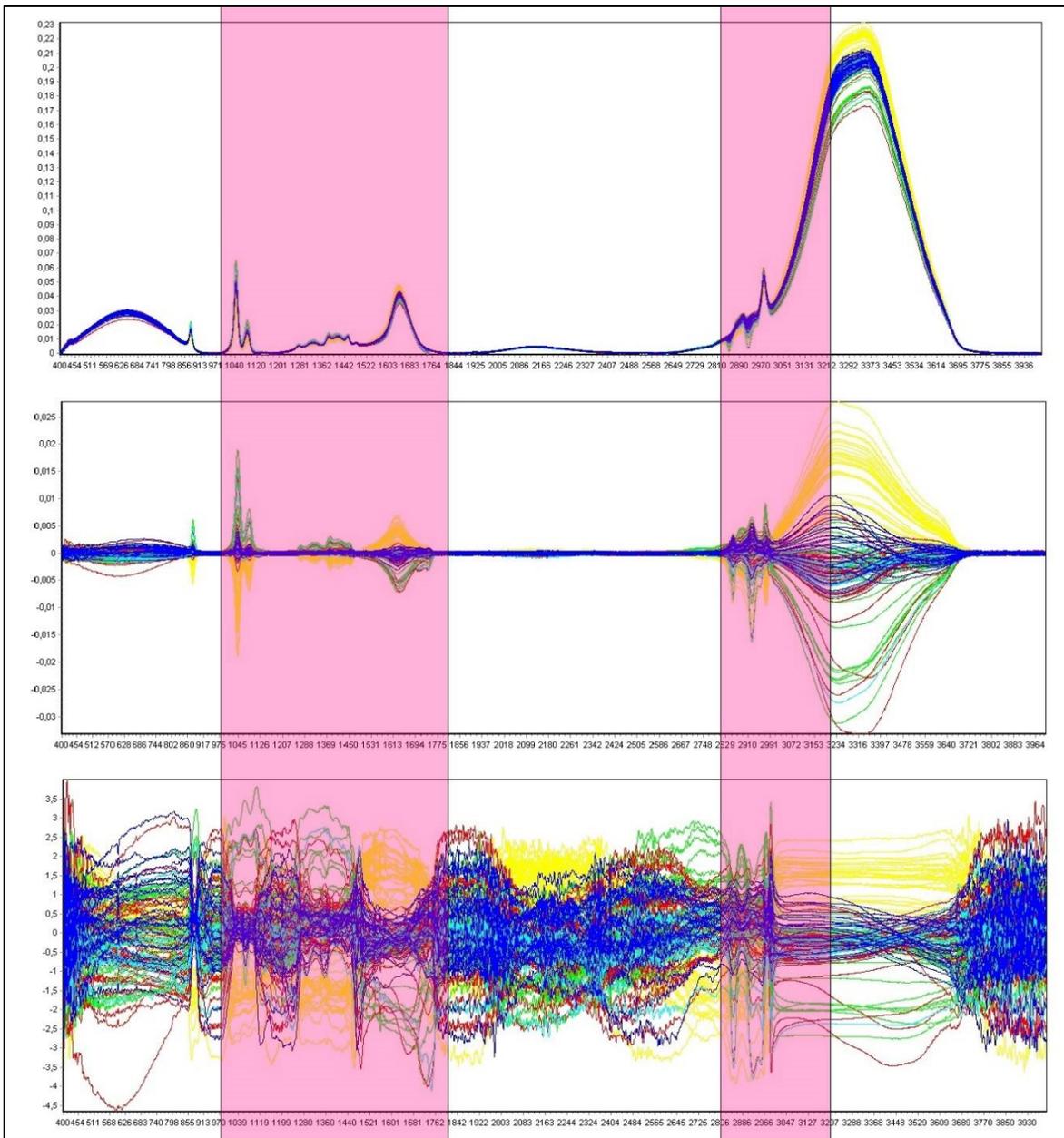


Figura 25.- Perfiles de los espectros FT-MIR con diferentes escalados, de arriba hacia abajo: sin escalado, centrado en la media y autoescalado. Intervalo optimo de trabajo resaltado en rosa.

Con este intervalo se probaron varios escalados y los modelos de PCA se resumen en las figuras 26, 27 y 28. Se muestra la varianza explicada por cada uno de los primeros 3 componentes principales con los diferentes tipos de escalados (ninguno, centrado en la media y autoescalado). También se incluyen los diagramas de scores PCA en el subespacio PC1-PC2 y la varianza acumulada al menos mayor al 94 %, para los 3 primeros componentes principales en todos los casos.

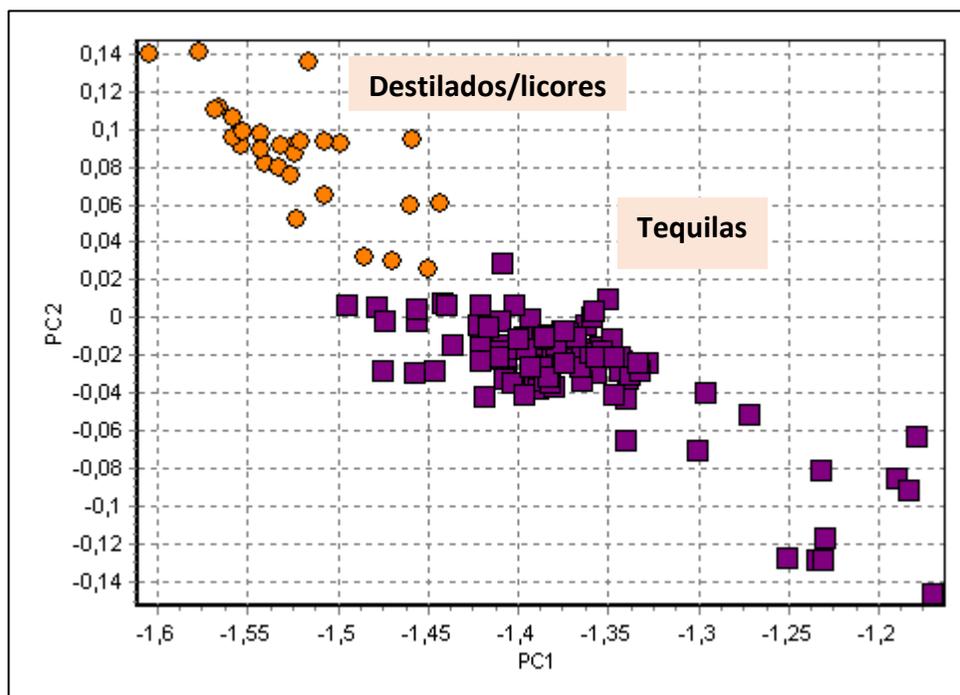


Figura 26.- Diagrama de scores en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Sin escalado. PC1: 99.8138 %, PC2: 0.1561 %, PC3: 0.0246 %.

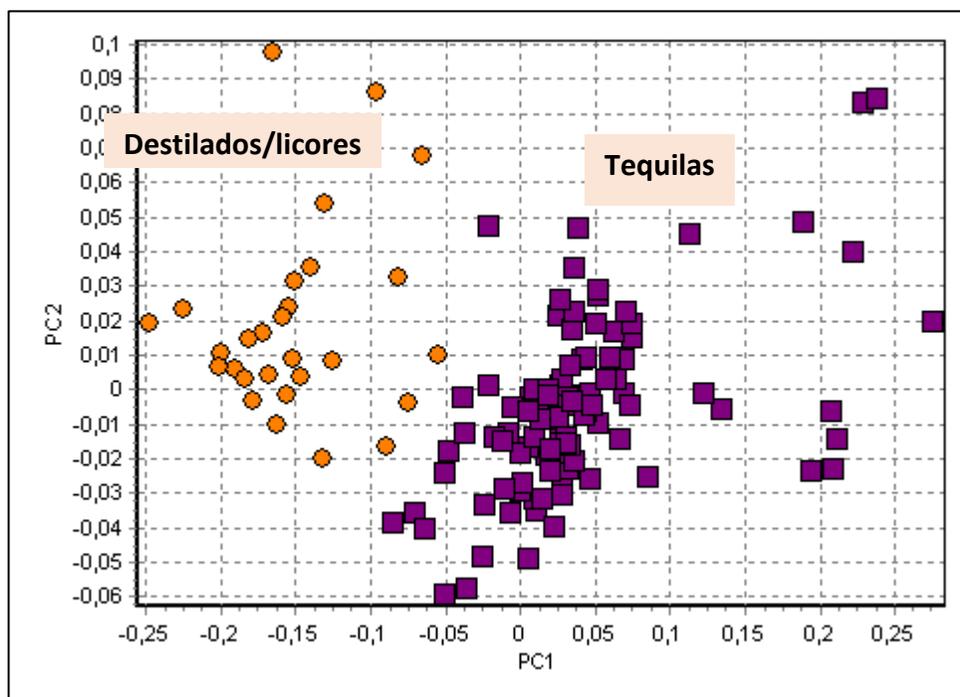


Figura 27.- Diagrama de scores en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media. PC1: 90.8301 %, PC2: 6.9695 %, PC3: 1.6127 %.

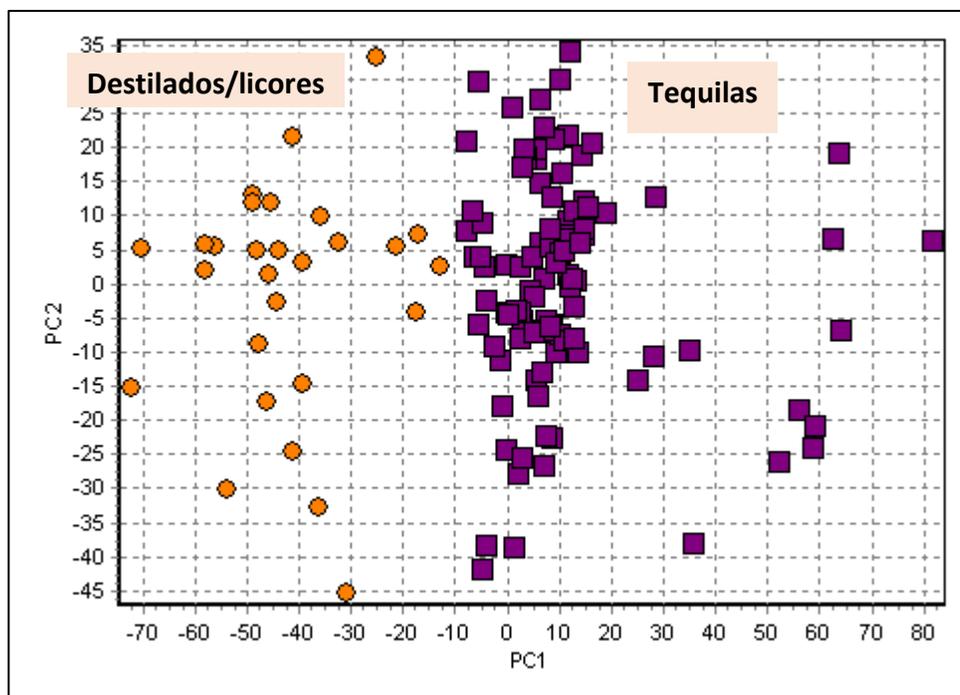


Figura 28.- Diagrama de scores en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 61.8607 %, PC2: 21.8087 %, PC3: 11.1533 %.

Con base en estos resultados, se seleccionó el autoescalado como el más adecuado, ya que en este caso se logró una separación muy clara lo largo de PC1 con 61.8607 % de varianza, siendo este componente el responsable de la separación entre los destilados/licores de agave del resto de tequilas. Así, al observar este eje, de -10 hacia el lado izquierdo, se ubica el grupo de los destilados/licores de agave y del lado derecho los tequilas en sus diferentes categorías y clases. El resultado obtenido se puede considerar satisfactorio.

6.1.1.3 Modelos para la diferenciación de licores/destilados vs tequila por clase

A continuación, se realizaron los modelos para la diferenciación de licores/destilados vs. tequila de acuerdo con su clase, con datos autoescalados, por ser el que gráficamente mostró una mejor separación entre los grupos.

En este caso, de forma general, se observa que la diferenciación entre el grupo destilados/licores de agave y el de tequila por clase aún es mejor. Básicamente es PC1 el responsable de dichas separaciones. Las figuras 29-32 resumen los gráficos de scores óptimos, los intervalos, escalados y varianza explicada de los tres primeros componentes principales para cada uno de los casos.

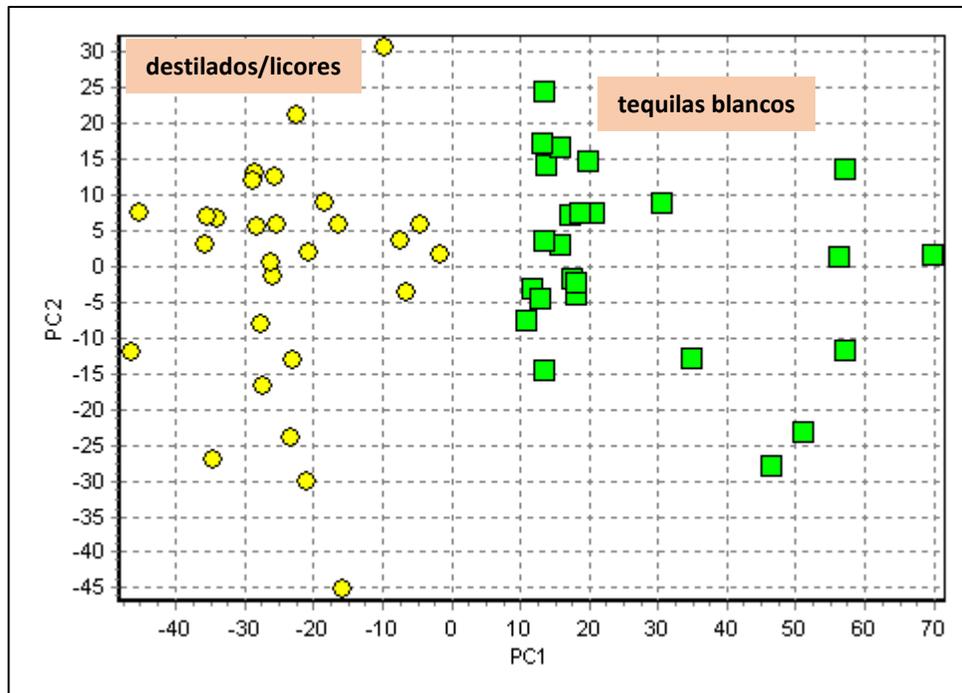


Figura 29.- Diagrama de scores destilados/licores vs. tequilas blancos en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 70.7163 %, PC2: 88.258 %, PC3: 96.9365 %.

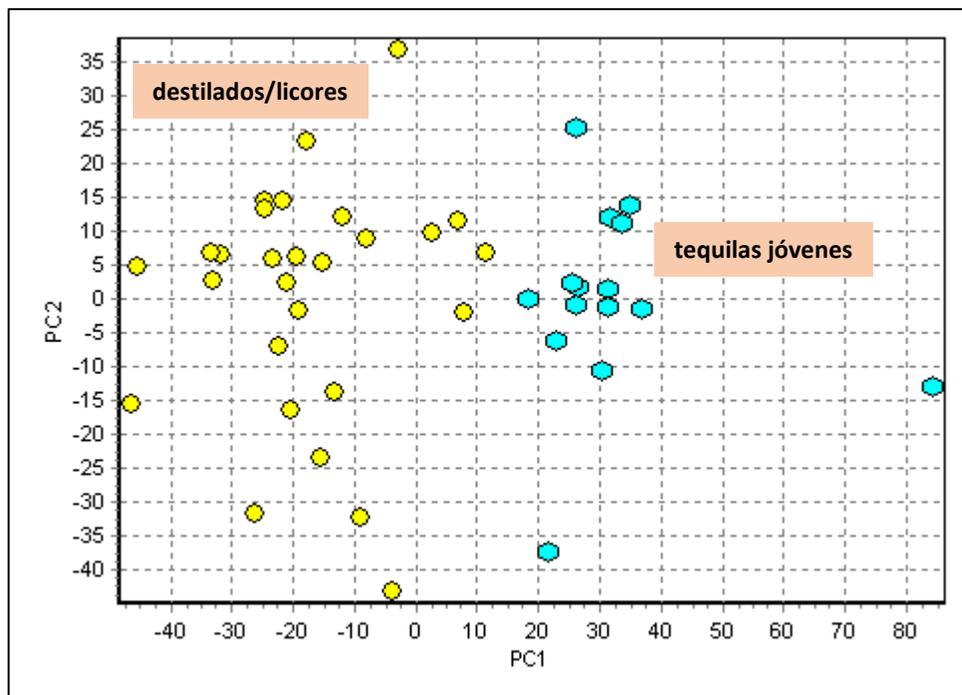


Figura 30.- Diagrama de scores destilados/licores vs. tequilas jóvenes en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 63.6495 %, PC2: 85.6239 %, PC3: 95.4386 %.

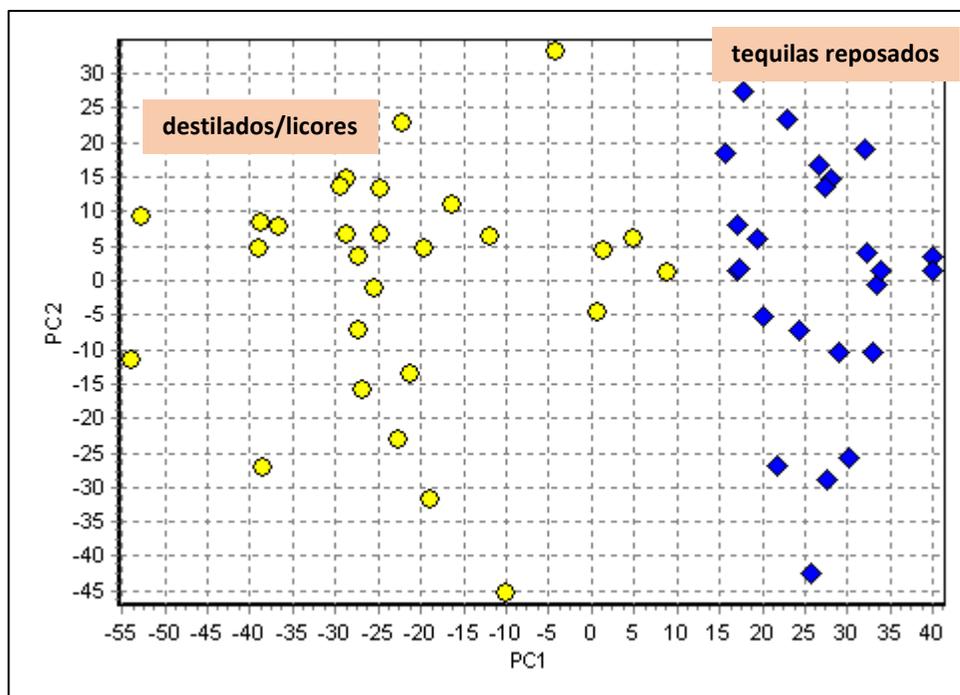


Figura 31.- Diagrama de scores destilados/licores vs. tequilas reposados en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 62.0827 %, PC2: 86.0065 %, PC3: 95.2580 %.

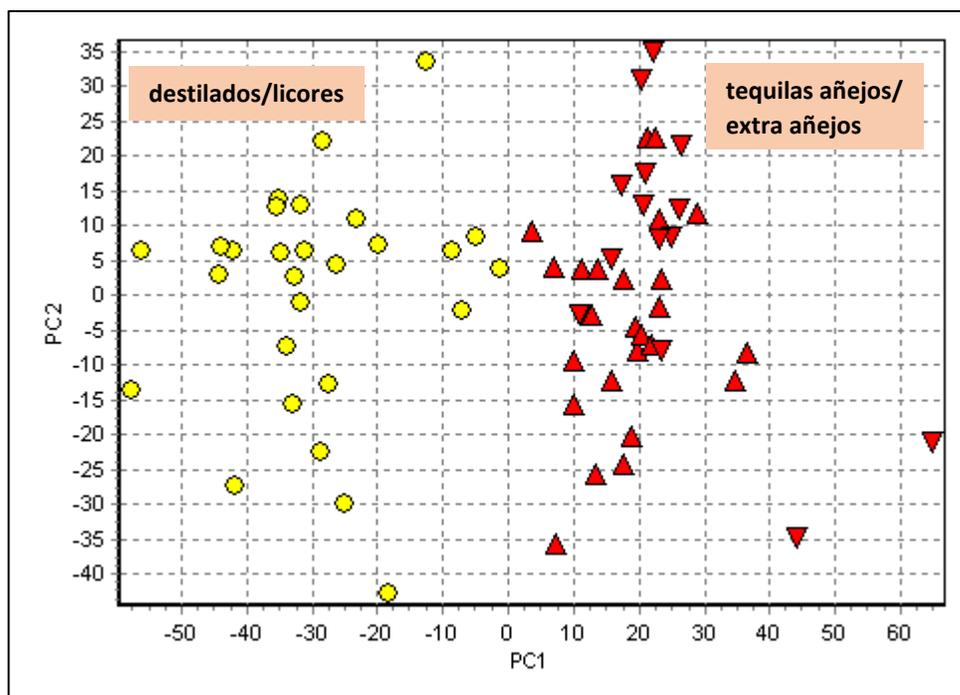


Figura 32.- Diagrama de scores destilados/licores vs. tequilas añejos/extra añejos en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Autoescalado. PC1: 62.6431 %, PC2: 84.8454 %, PC3: 95.0718 %.

6.1.2 Análisis Cluster (CA)

6.1.2.1 Modelos de CA para la diferenciación de tequilas vs licores/destilados

Para el análisis cluster el mejor modelo se obtuvo usando los intervalos de 980-1800 y 2800-3200 cm^{-1} , con datos autoescalados, aplicando algoritmo de Ward y distancia Euclídea cuadrada. El dendrograma de la figura 33 muestra dos agrupaciones diferentes. En la primera (1, en la figura) se observan contenidos los destilados/licores de agave en amarillo y mientras que en la segunda (2, en la figura) se ubican en color púrpura principalmente los tequilas entre los cuales se contienen 3 muestras de primer grupo, por lo que se puede considerar que este modelo es satisfactorio para lograr la diferenciación entre ambos grupos.

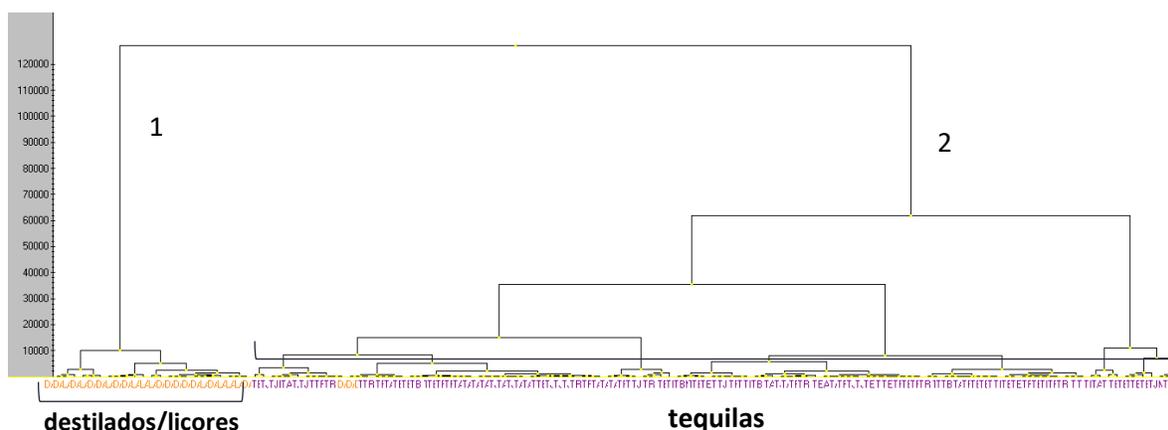


Figura 33.- Dendrograma para el conjunto de destilados/licores y tequilas, algoritmo de Ward y distancia Euclídea cuadrada. Autoescalado, intervalos: 980-1800 cm^{-1} y 2800-3200 cm^{-1} .

6.1.2.2 Modelos de CA para la diferenciación de licores/destilados vs tequila por clase

De manera análoga al análisis PCA, se realizó el análisis cluster para la diferenciación de destilados/licores y tequila por clase en las mismas condiciones del CA previo.

Las figuras 34-37 muestran los dendrogramas de los análisis cluster para la diferenciación de los destilados/licores de agave con cada una de las clases de tequila. Se observa que todas las clases de tequila están separadas de los destilados/licores de agave. En todos los casos, se observan dos agrupaciones. El primer grupo está conformado sólo por todos los destilados/licores y en el segundo se agrupan los tequilas de la correspondiente clase. En cada dendrograma cada grupo contiene sólo un tipo de muestras a excepción del dendrograma de tequilas jóvenes donde algunas muestras destilados/licores se agrupan en

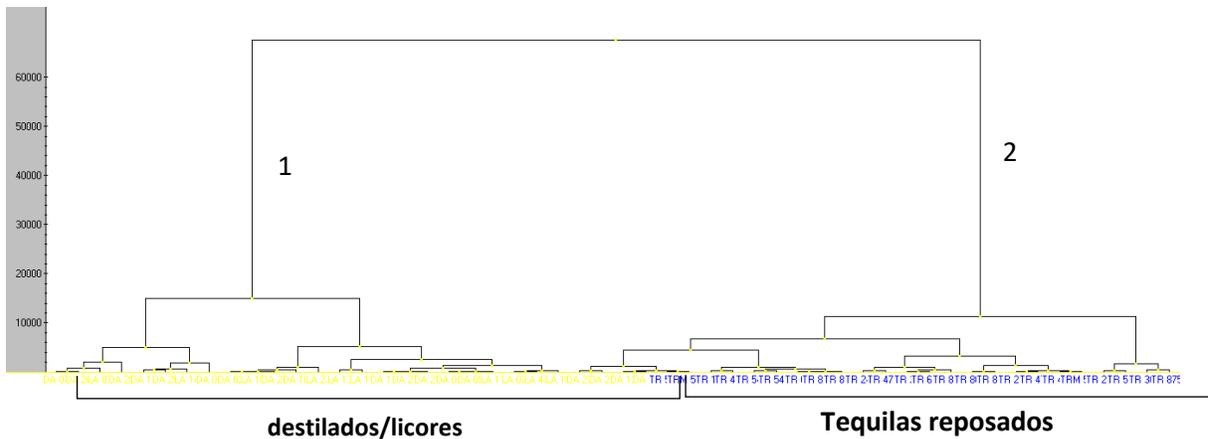


Figura 36.- Dendrograma para el conjunto de destilados/licores y tequilas reposados, algoritmo de Ward y distancia Euclídea cuadrada. Autoescalado, intervalos: 980-1800 cm^{-1} y 2800-3200 cm^{-1} .

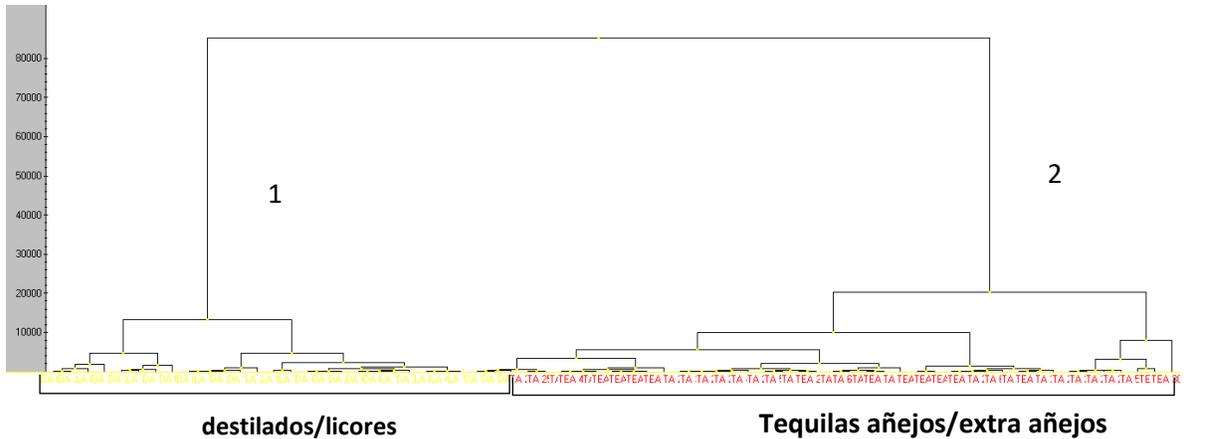


Figura 37.- Dendrograma para el conjunto de destilados/licores y tequilas añejos/extra añejos, algoritmo de Ward y distancia Euclídea cuadrada. Autoescalado, intervalos: 980-1800 cm^{-1} y 2800-3200 cm^{-1} .

Por tanto, las técnicas exploratorias de PCA y Cluster permiten realizar un reconocimiento de los grupos de muestras y las regiones espectrales que los determinan. Ambos análisis nos revelan información complementaria y coherente.

PCA permitió detectar las muestras anómalas y quitarlas del modelo. Encamino el análisis a realizarlo enfrentando de manera individual a cada clase de tequila con el conjunto de destilados/licores de agave.

CA mostró que la diferencia entre tequilas y destilados/licores de agave es grande; la distancia de los dendrogramas es clave al analizar este punto.

6.2 Técnicas supervisadas.

Con la información obtenida de los análisis exploratorios previos, se puede proseguir con mayor conocimiento y confianza, con las técnicas supervisadas. Ya sabemos que, en efecto, las muestras de tequilas son diferentes a las de los destilados/licores de agave. Ahora es momento de avanzar en su estudio con las técnicas de clasificación. En el presente estudio dichas técnicas serán Curvas de Potencia (CP) y Support Vector Machines (SVM). CP tiene las ventajas de ser una herramienta muy sencilla y de clasificar las muestras asignando una probabilidad de pertenencia a los diferentes grupos, siendo el científico el que toma la decisión final. SVM permite abordar problemas más complejos donde los solapamientos de los conjuntos de muestras son elevados y, por tanto, pudieran no separarse empleando otras técnicas en las que se generan hiperplanos (es decir, se trabaja dimensionalidad reducida). Por el contrario SVM considera más dimensiones que las de partida.

6.2.1 Curvas de potencia

En este caso no se presenta un modelo que involucre todas las clases de destilados/licores y tequilas, pues no fue posible lograr su diferenciación.

6.2.1.1 Modelos de CP para la diferenciación de licores/destilados vs tequila por clase

Para el análisis por curvas de potencia fue necesario asignar 2 grupos de muestras, el grupo mayoritario que fue el de calibración, con el cual el algoritmo construye el modelo, y el segundo grupo minoritario, el de validación con el cual se pone a prueba el modelo. Se tomaron 114 muestras para el grupo de calibración y 17 más para el de validación.

Los resultados óptimos, presentados en las figuras 38-41, se obtuvieron con los intervalos $980 - 1800 \text{ cm}^{-1}$ y $2800 - 3200 \text{ cm}^{-1}$, con centrado en la media. Las tablas 2-5 muestran los valores de probabilidad de pertenencia de las muestras de validación para cada uno de los análisis de CP.

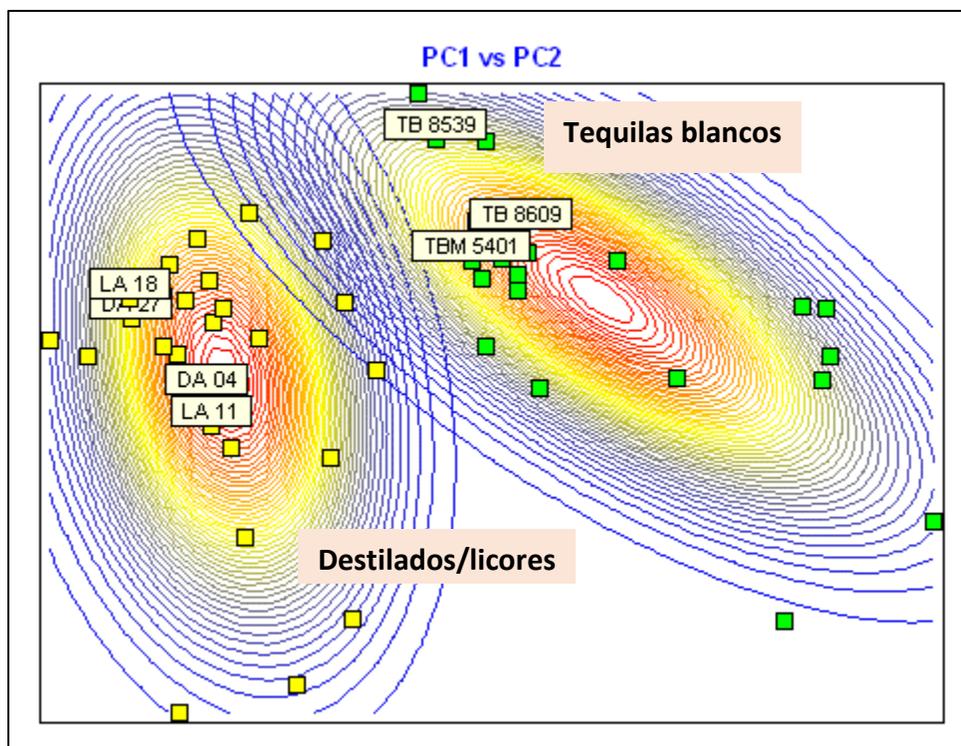


Figura 38.- Curvas de potencia para destilados/licores vs. tequilas blancos en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media.

Tabla 2. Probabilidades de pertenencia de las muestras de validación para destilados/licores vs tequilas blancos

Muestra	% destilados	% blancos
DA 04	96.35	0
DA 27	52.89	0
LA 11	90.48	0
LA 18	49.67	0
TB 8539	0	38.55
TB 8609	0	82.99
TBM 5401	0	65.45

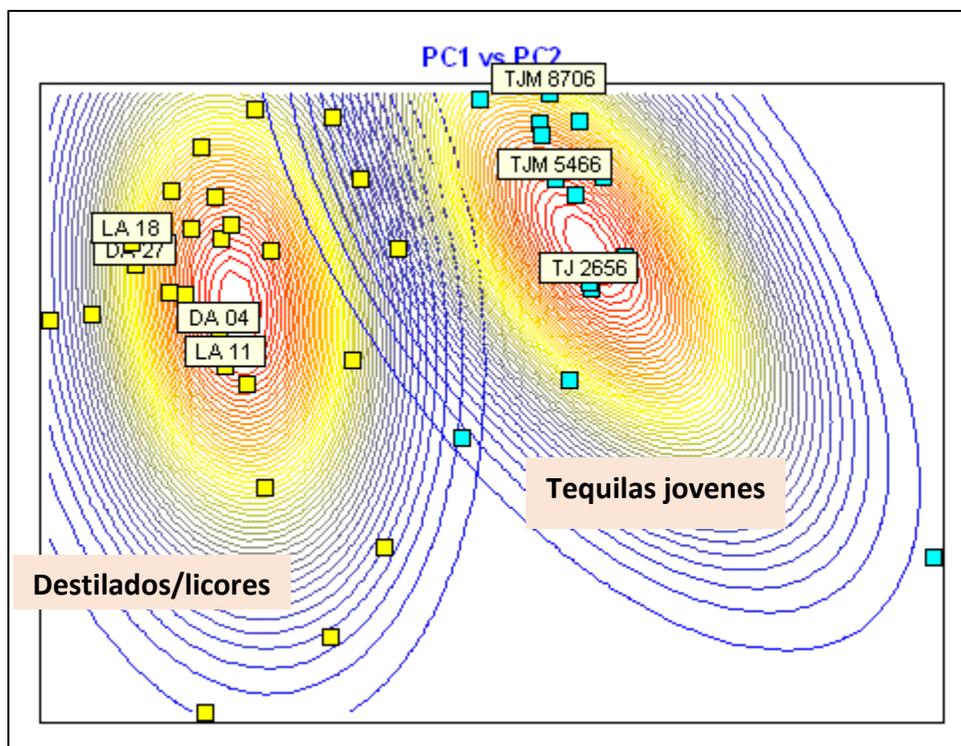


Figura 39.- Curvas de potencia para destilados/licores vs. tequilas jóvenes en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media.

Tabla 3. Probabilidades de pertenencia de las muestras de validación para destilados/licores vs tequilas jóvenes

Muestra	% destilados	% jóvenes
DA 04	95.99	0
DA 27	52.91	0
LA 11	90.32	0
LA 18	49.72	0
TJ 2656	0	95.86
TJM 5466	0	90.46
TJM 8706	0	53.65

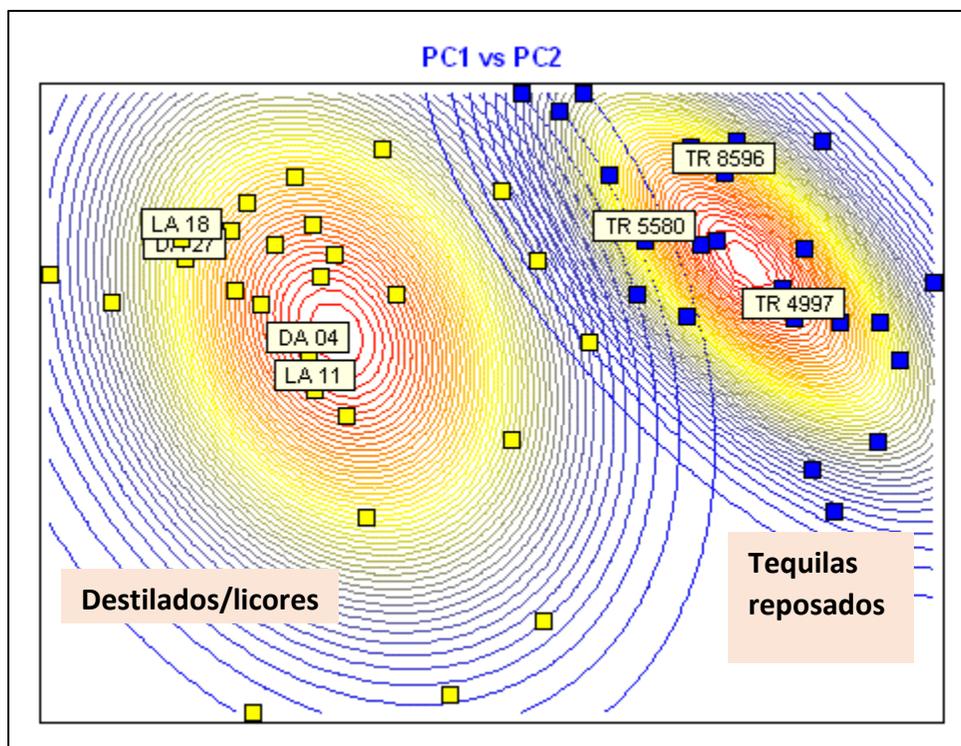


Figura 40.- Curvas de potencia para destilados/licores vs. tequilas reposados en los intervalos $980-1800\text{ cm}^{-1}$ y $2800-3200\text{ cm}^{-1}$, Centrado en la media.

Tabla 4. Probabilidades de pertenencia de las muestras de validación para destilados/licores vs tequilas reposados.

Muestra	% destilados	% reposados
DA 04	96.5	0
DA 27	52.93	0
LA 11	90.54	0
LA 18	49.78	0
TR 4997	0	89.4
TR 5580	3.49	59.98
TR 8596	0	69.44

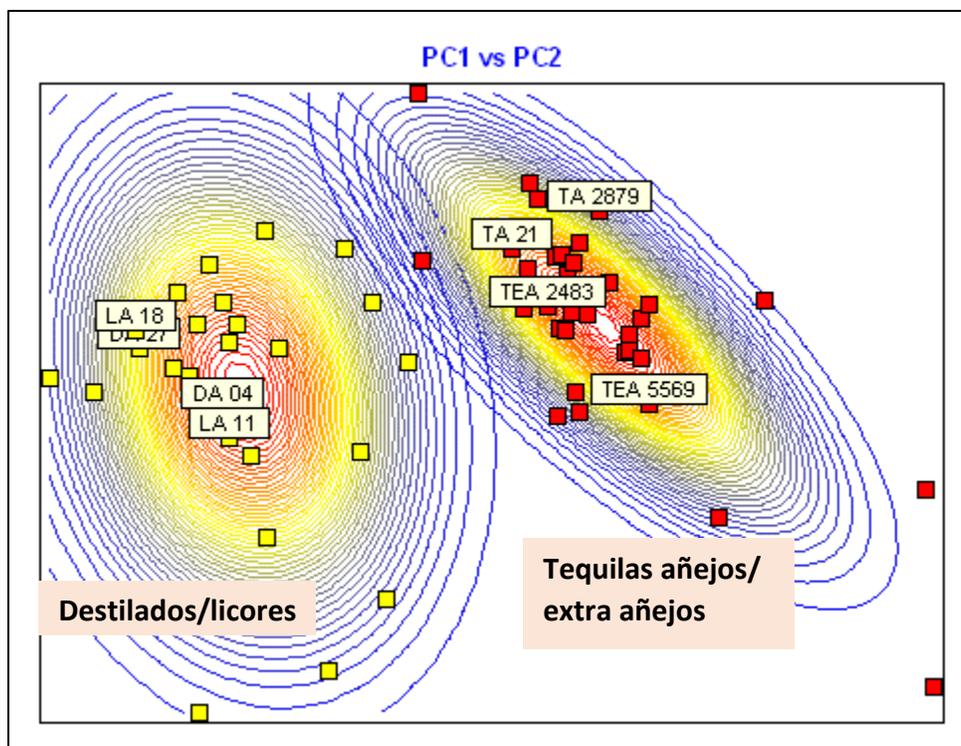


Figura 41.- Curvas de potencia para destilados/licores vs. tequilas añejos/extra añejos en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , Centrado en la media

Tabla 5. Probabilidades de pertenencia de las muestras de validación para destilados/licores vs tequilas añejos/extra añejos.

Muestra	% destilados	% añejos
DA 04	96.37	0
DA 27	52.88	0
LA 11	90.48	0
LA 18	49.79	0
TA 21	0	68.64
TA 2879	0	22.69
TEA 2483	0	81.39
TEA 5569	0	70.73

En las tablas 2-5 se observa claramente que todas las muestras de validación se asignan a sus respectivos grupos con porcentajes grandes de probabilidad de pertenencia. Esto es un indicador de que los modelos de CP son satisfactorios. Ello a pesar de que gráficamente se aprecia un ligero solapamiento entre las regiones del espacio de trabajo asignadas a cada clase.

Para poder evaluar la capacidad de predicción de los modelos, se hizo uso del método de la validación cruzada [Otto, 2017]. En este método cada una de las muestras de calibración se asigna en una ocasión como muestra de validación. Se realizó el conteo de muestras que

fueron asignadas correctamente en cada uno de los análisis y se calculó el porcentaje de muestras que fueron asignadas correctamente a su respectivo grupo. Este dato de porcentaje de asignación está relacionado con la capacidad de predicción; entre mayor sea este valor, es menor la probabilidad del error de que una muestra sea asignada incorrectamente a su grupo. El análisis por curvas de potencia nos da la probabilidad de pertenencia de cada muestra de validación a un determinado grupo. Este valor nos permite evaluar la confiabilidad de la regla de clasificación.

La tabla 6 muestra los resultados del análisis de capacidad de predicción para cada uno de los análisis de curvas de potencia.

Tabla 6. Análisis de capacidad de predicción para CP.

Muestras	Capacidad de predicción	Porcentaje de probabilidad de pertenencia a su grupo.
Destilados-Blancos	96.22 %	51.86 %
Destilados-Jóvenes	90.69 %	56.04 %
Destilados-Reposados	98.07 %	49.17 %
Destilados-Añejos/Extra añejos	94.02 %	56.74 %

Los resultados de curvas de potencia se consideran satisfactorios, con una capacidad de predicción del método que es mayor al 90 % en todos los casos. Por otra parte, el porcentaje de probabilidad de pertenencia al grupo tiene valores que rondan entre 49-57 % de probabilidad; esto significa que los grupos formados no están lo suficientemente cohesionados entre sí para que todas las muestras tengan valores altos de probabilidad de pertenencia. Esto se atribuye principalmente a diferencias físicas y químicas de las muestras, diferencias tales como periodos de reposo, abocamientos o forma de producción del producto (artesanal o industrial). Sin embargo estos valores son lo suficientemente grandes para que la regla de clasificación asigne a más del 90 % de las muestras en su grupo correcto.

6.2.2 Clasificación mediante vectores soporte (*Support Vector Machines, SVM*)

6.2.2.1 Modelos de SVM para la discriminación de tequilas vs licores/destilados

Support Vector Machines es un método que requiere de una codificación específica de la matriz de datos. Al igual que en curvas de potencia, SVM requiere de un conjunto de muestras de calibración y otro de validación. La regla de clasificación asigna cada una de las muestras de validación dependiendo de su ubicación en los hiperespacios calculados.

El primer modelo fue realizar un análisis SVM que involucrara todas las muestras de destilados/licores y tequilas (sin considerar sus clases). Es necesario nuevamente hacer notar que el número de muestras de tequilas es superior al de licores/destilados; el primero es de 28, en tanto que el del segundo es de 103. No obstante, se intentó el modelo con resultados satisfactorios. Para seleccionar las muestras de validación se hizo uso de los gráficos de PCA. Las muestras de validación fueron seleccionadas de manera que los scores de éstas estuviesen distribuidas en todo el subespacio del diagrama de scores.

La tabla 7 presenta el conjunto de muestras de validación seleccionadas.

Tabla 7. Selección de muestras de validación.

Muestra	Clave
Destilados de agave	DA04, DA27
Licores de agave	LA11, LA18
Tequilas Blancos	TB8539, TB8609, TBM5401
Tequilas Jóvenes	TJ2656, TJM5466, TJM8706
Tequilas Reposados	TR4997, TR5580, TR8596
Tequilas Añejos	TA21, TA2879
Tequilas Extra añejos	TEA2483, TEA5569

El análisis óptimo, mostrado en la figura 42, se obtuvo en los intervalos 980-1800 cm^{-1} y 2800-3200 cm^{-1} , con centrado en la media, Kernel lineal y $C=50$. Recuérdese que C es el factor de penalización para errores cometidos en la clasificación. La tabla 8 muestra los valores de asignación de las muestras de validación; la codificación 1 para destilados/licores de agave y 2 para tequilas en general.

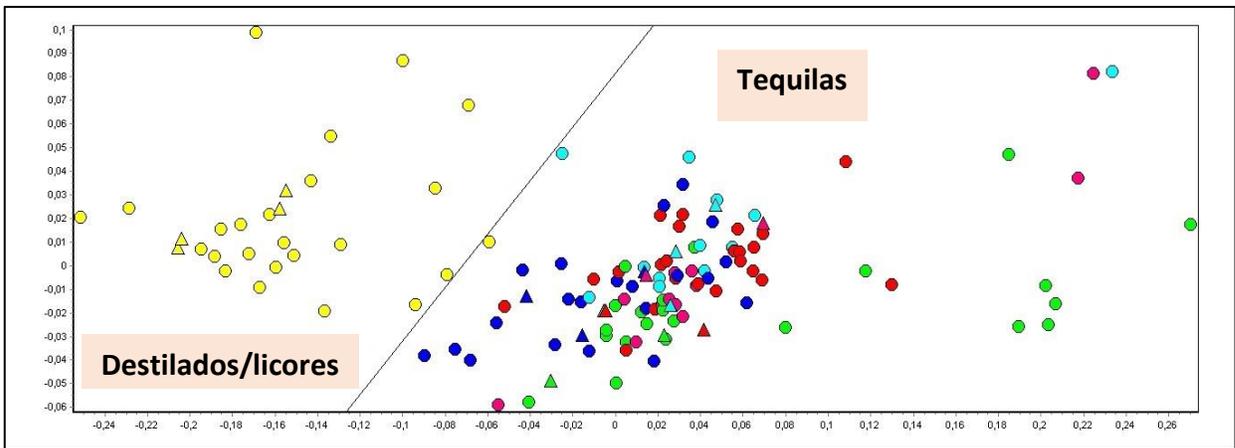


Figura 42.- Análisis óptimo para SVM considerando Destilados/licores y tequilas en todas sus clases en los intervalos $980-1800\text{ cm}^{-1}$, $2800-3200\text{ cm}^{-1}$, Kernel lineal, $C=50$, 10000 iteraciones.

Tabla 8. Asignación de las muestras de validación para destilados/licores vs tequilas vía SVM.

kernel:lineal (mean center)	
Muestra	Destilados vs resto
DA 04	1
DA 27	1
LA 11	1
LA 18	1
TA 21	2
TA 2879	2
TB 8539	2
TB 8609	2
TBM 5401	2
TEA 2483	2
TEA 5569	2
TJ 2656	2
TJM 5466	2
TJM 8706	2
TR 4997	2
TR 5580	2
TR 8596	2

6.2.2.2 Modelos de SVM para la discriminación de licores/destilados vs tequila por clase.

Al igual que en los casos anteriores, se realizaron los análisis SVM para la discriminar el conjunto de destilados/licores de agave de cada una de las diferentes clases de tequila.

Las figuras 43-46 muestran los gráficos SVM óptimos. Las tablas 9-12 muestran los valores de asignación de las muestras de validación; la codificación 1 para destilados/licores de agave y 2 para tequilas en general. Éstos se obtuvieron en los intervalos 980-1800 y 2800-3200 cm^{-1} . El escalado óptimo fue centrado en la media. Adicional a esto se probaron diferentes Kernel: lineal, gaussiano y polinomial. Los tres tipos de Kernel dan excelente resultado; sin embargo, se elige como óptimo el Kernel lineal por su mayor simplicidad.

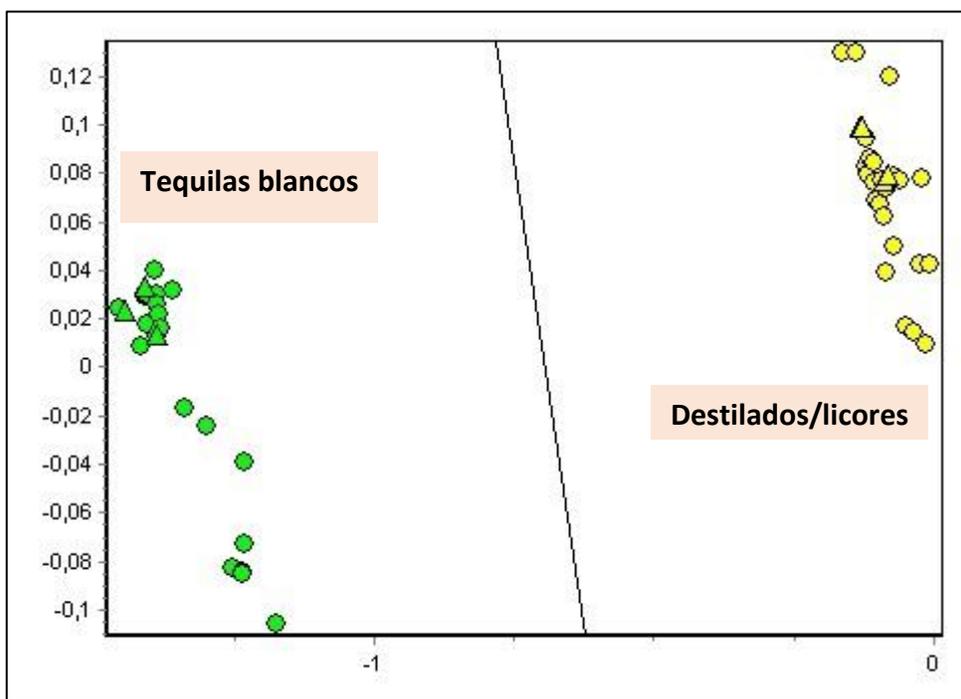


Figura 43.- Gráfico SVM para destilados/licores vs. tequilas blancos, Centrado en la media en los intervalos 980-1800 cm^{-1} , 2800-3200 cm^{-1} , Kernel lineal, $C=50$, 10000 iteraciones.

Tabla 9. Asignación de las muestras de validación para destilados/licores vs tequilas blancos vía SVM.

kernel:lineal (mean center)	
Muestra	Destilados vs resto
DA 04	1
DA 27	1
LA 11	1
LA 18	1
TB 8539	2
TB 8609	2
TBM 5401	2

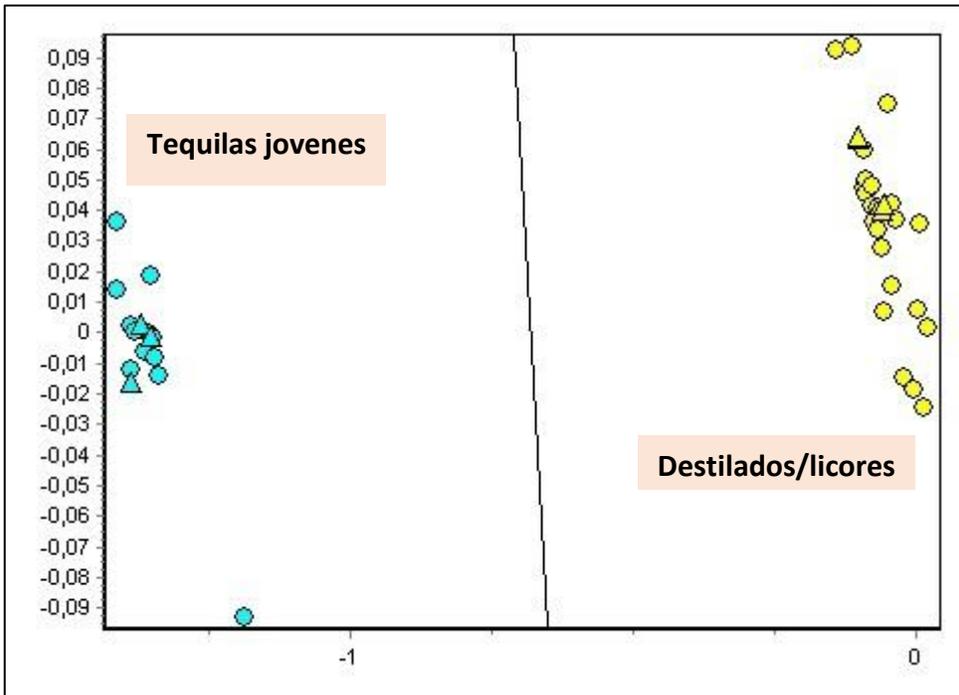


Figura 44.- Grafico SVM para destilados/licores vs. tequilas jóvenes, Centrado en la media en los intervalos $980-1800\text{ cm}^{-1}$, $2800-3200\text{ cm}^{-1}$, Kernel lineal, $C=50$, 10000 iteraciones.

Tabla 10. Asignación de las muestras de validación para destilados/licores vs tequilas jóvenes vía SVM.

kernel:lineal (mean center)	
Muestra	Destilados vs resto
DA 04	1
DA 27	1
LA 11	1
LA 18	1
TJ 2656	2
TJM 5466	2
TJM 8706	2

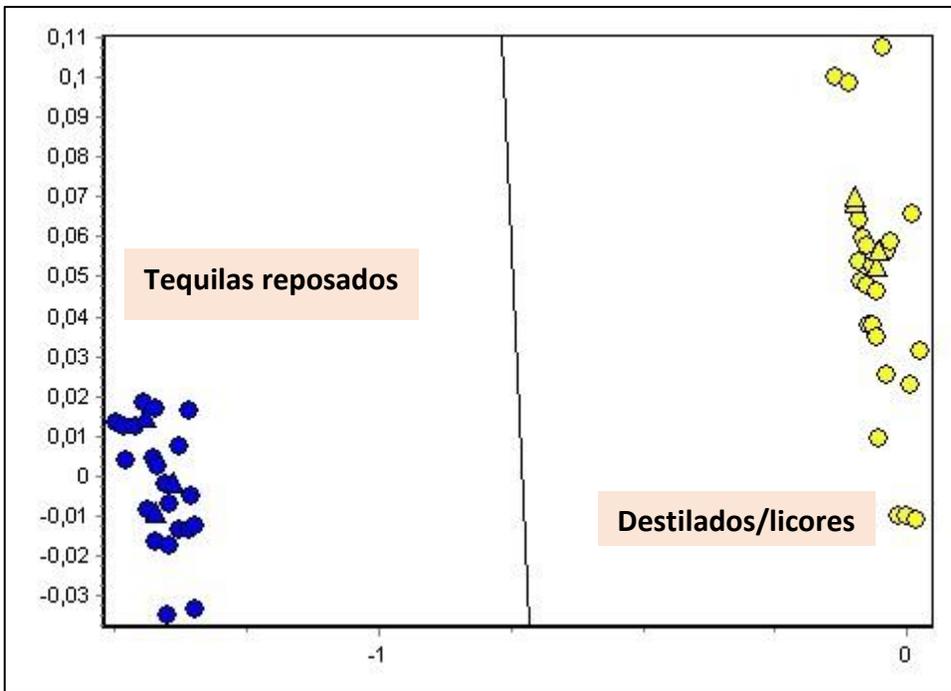


Figura 45.- Grafico SVM para destilados/licores vs. tequilas reposados, Centrado en la media en los intervalos $980-1800\text{ cm}^{-1}$, $2800-3200\text{ cm}^{-1}$, Kernel lineal, $C=50$, 10000 iteraciones.

Tabla 11. Asignación de las muestras de validación para destilados/licores vs tequilas reposados vía SVM.

kernel:lineal (mean center)	
Muestra	Destilados vs resto
DA 04	1
DA 27	1
LA 11	1
LA 18	1
TR 4997	2
TR 5580	2
TR 8596	2

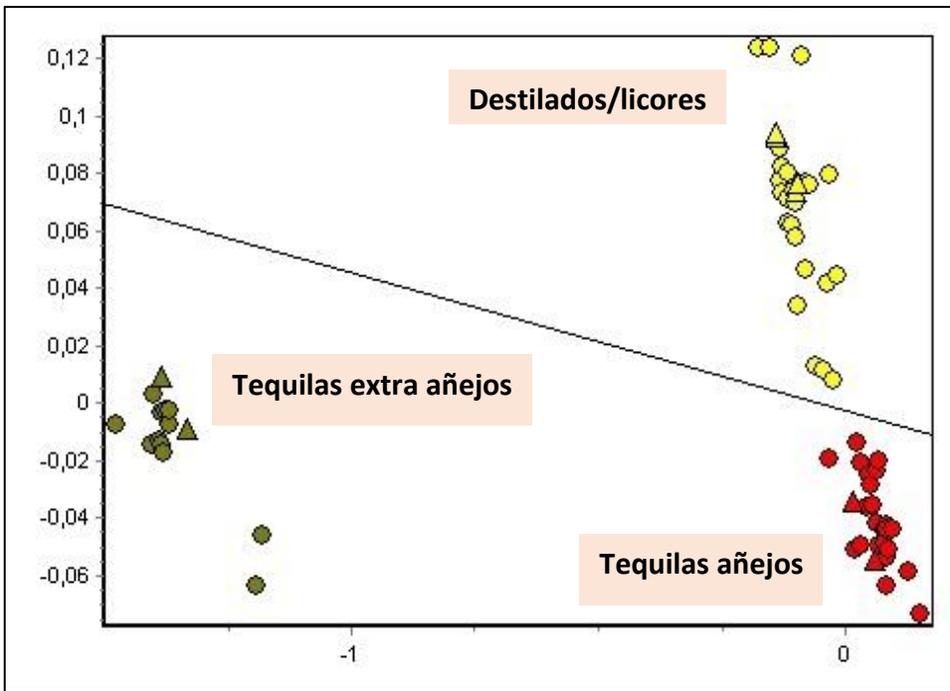


Figura 46.- Gráfico SVM para destilados/licores vs. tequilas añejos/extra añejos, Centrado en la media en los intervalos 980-1800 cm^{-1} , 2800-3200 cm^{-1} , Kernel lineal, $C=50$, 10000 iteraciones.

Tabla 12. Asignación de las muestras de validación para destilados/licores vs tequilas añejos/extra añejos vía SVM.

kernel:lineal (mean center)	
Muestra	Destilados vs resto
DA 04	1
DA 27	1
LA 11	1
LA 18	1
TA 21	2
TA 2879	2
TEA 2483	2
TEA 5569	2

De las figuras anteriores podemos observar que se logra una muy buena separación de las muestras; sin embargo, el resultado más interesante se puede observar en el análisis de SVM de destilados/licores de agave contra tequilas añejos/extra añejos (figura 46). Intencionalmente se asignó el color rojo para las muestras de tequilas añejos y el color verde olivo a los tequilas extra añejos. La regla de clasificación es capaz de diferenciar los destilados/licores de agave de las muestras de tequila, pero además el modelo separa las muestras de tequilas añejos de los tequilas extra añejos, sin combinar ninguna de estas clases de tequila con los destilados/licores de agave. Este es una clara muestra de las

diferencias existentes entre muestras y del poder resolutivo de los métodos quimiométricos.

Para SVM también se realizó el análisis de capacidad de reconocimiento y capacidad de predicción por el método de las navajas [Otto, 2017]. El resultado de este análisis se presenta en la tabla 13.

Tabla 13. Análisis de capacidades de reconocimiento y predicción para SVM.

Muestras	Capacidad de reconocimiento	Capacidad de predicción
Destilados-Blancos	100 %	100 %
Destilados-Jóvenes	100 %	100 %
Destilados-Reposados	100 %	100 %
Destilados-Añejos/Extra añejos	100 %	100 %

En SVM se observa un resultado excelente en cuanto al análisis de las capacidades de reconocimiento y de predicción. Los resultados de asignación de las muestras de calibración y validación son en todos los casos correspondientes a su grupo de pertenencia. No se puede esperar un mejor resultado que un modelo en el que sus valores de reconocimiento y de predicción sean muy similares; se puede entonces afirmar que los modelos generados son muy confiables y robustos ya que no depende del conjunto de muestras seleccionadas.

7.0 CONCLUSIONES

- Se logró la diferenciación de los destilados/licores de agave por las técnicas no supervisadas de Análisis de Componentes Principales (PCA) y Análisis Cluster (CA) con resultados similares. Las condiciones óptimas fueron: intervalos de trabajo de 980-1800 cm^{-1} y 2800-3200 cm^{-1} , con autoescalado.
- Se logró la diferenciación y clasificación de los destilados/licores de agave por las técnicas supervisadas Curvas de Potencia (CP) y Support Vector Machines (SVM). Las condiciones óptimas fueron: intervalo de trabajo de 980-1800 cm^{-1} y 2800-3200 cm^{-1} , con centrado en la media. Las capacidades de reconocimiento y de predicción fueron mejores con SVM que con CP. Además, SVM no sólo fue capaz de discriminar entre destilados/licores de agave, tequilas añejos, sino también discrimina a la clase tequilas extra añejos. Esto no era propósito del trabajo pero sirve de evidencia para demostrar el poder resolutivo de las técnicas quimiométricas.

8.0 REFERENCIAS

1. García A. J, (julio-septiembre, 2007). Los agaves de México. *Ciencias*. Recuperado de:
<http://www.revistaciencias.unam.mx/images/stories/Articles/87/02/Los%20agaves%20de%20Mexico.pdf>
2. Parsons J. R, (18 de febrero de 2012). Los agaves en la economía tradicional. *La Jornada del Campo*. Recuperado de:
<http://www.jornada.com.mx/2012/02/18/cam-economia.html>
3. García A. J, (18 de febrero de 2012). México, país de Magueyes. *La Jornada del Campo*. Recuperado de: <http://www.jornada.com.mx/2012/02/18/cam-pais.html>
4. Norma Oficial Mexicana NOM-006-SCFI-2012, Bebidas alcohólicas-Tequila-Especificaciones, Diario Oficial de la Federación, Ciudad de México, México, 13 de diciembre de 2012.
5. Norma Oficial Mexicana NOM-159-SCFI-2004, Bebidas alcohólicas-Sotol-Especificaciones y Métodos de Prueba, Diario Oficial de la Federación, Ciudad de México, México, 16 de junio de 2004.
6. Norma Oficial Mexicana NOM-168-SCFI-2004, Bebidas alcohólicas-Bacanora-Especificaciones de elaboración, envasado y etiquetado, Diario Oficial de la Federación, Ciudad de México, 14 de diciembre de 2005.
7. Norma Oficial Mexicana NOM-070-SCFI-2016, Bebidas alcohólicas-Mezcal-Especificaciones, Diario Oficial de la Federación, Ciudad de México, México, 23 de febrero de 2017.
8. Proyecto de Norma Oficial Mexicana PROY-NOM-186-SCFI-2012, Bebidas alcohólicas-Bebidas alcohólicas elaboradas a partir de agave-Especificaciones, métodos de prueba e información comercial, Diario Oficial de la Federación, Ciudad de México, México, 05 de julio de 2012.
9. Williams, I., (2015), *Tequila A Global History*, Londres, Reino Unido: Reaktion Books Ltd.
10. Yurkanis, P., (2008), *Química Orgánica*, Naucalpan de Juárez, México: Pearson Educación.
11. Christian, G., (2009), *Química Analítica*, CDMX, México: McGraw-Hill.
12. Levine, I., (1991), *Fisicoquímica*, Madrid, España: McGraw-Hill.
13. Mancilla-Margalli, N., A., y López, M., G. (2002). Generation of Maillard Compounds from Inulin during the Thermal Processing of Agave tequilana Weber Var. Azul. *Journal of Agricultural and Food Chemistry*, 50(4), 806-812.
14. Harris, D., (2010), *Análisis químico cuantitativo*, Barcelona, España: Editorial Reverté.
15. Peña, D., (2002), *Análisis de datos multivariantes*, Madrid, España: McGraw-Hill.
16. Massart, D., L., Vandeginste, B., G., M., Deming, S., N., Michotte, Y. y Kaufman, L., (2003), *Chemometrics: a textbook*, Ámsterdam, Países Bajos: El Sevier.

17. Otto, M., (2017), *Chemometrics: Statistics and Computer Application in Analytical Chemistry*, Weinheim, Alemania: Wiley-VCH Verlag GmbH & Co. KGaA
18. Everitte, B. S., Dunn, G., (2001), *Applied Multivariate Data Analysis*, Londres, Gran Bretaña: John Wiley & Sons, Ltd.
19. Miller, J. N., Miller, J. C., (2002), *Estadística y quimiometría para química analítica*, Madrid, España: Prentice Hall.
20. Kulkarni, S., Harman, G., (2011), *An Elementary Introduction to Statistical Learning Theory*, Nueva Jersey, EUA: John Wiley & Sons, Inc.

9.0 REFERENCIAS DE SOFTWARE.

1. Datan (Versión: 5.0.6.167) [Software]. (2013). MultiD Analyses AB. Obtenido de: www.multid.se
2. Spectrum (Versión: 10.5.2.636) [Software]. (2015). Perkin Elmer, Inc.
3. Microsoft Excel (Versión: 15.0.5197.1000) [Software]. (2013). Microsoft corporation.

10.0 REFERENCIAS DE FIGURAS.

1. García, T., Rico, V., (2016 a). Distintas variedades de agave. [Figura].
2. García, T., Rico, V., (2016 b). Distribución de agaves en la república mexicana. [Figura].
3. Matadamas, D., (2016). Mayáhuel, diosa del maguey. Códice Borbónico, lám. 8. [Figura]. Recuperado de: <https://arqueologiamexicana.mx/mexico-antiguo/las-deidades-del-pulque>
4. CRT, (2019 a). Producción Total: Tequila y Tequila 100 %. [Figura]. Recuperado de: <https://www.crt.org.mx/EstadisticasCRTweb/>
5. Adaptado de CRT, (2019 b). Denominación de origen del tequila. [Figura]. Recuperado de: https://www.crt.org.mx/index.php?option=com_wrapper&view=wrapper&Itemid=408
6. García, T., Rico, V., (2016 c). Agave tequilana Weber variedad azul con rizomas expuestos. [Figura].
7. García, T., Rico, V., (2016 d). Rasurado de piña de agave. [Figura].
8. García, T., Rico, V., (2016 e). Horno de mampostería. [Figura].
9. García, T., Rico, V., (2016 f). Tahona. [Figura].
10. García, T., Rico, V., (2016 g). Tanques de fermentación de acero inoxidable. [Figura].
11. García, T., Rico, V., (2016 h). Alambiques de cobre y contenedores de acero inoxidable de tequila ordinario y tequila. [Figura].

12. García, T., Rico, V., (2016 i). Barricas. [Figura].
13. Adaptado de Excel® (Professional Plus). [Software]. 2013. Microsoft Corporation®.
14. Otto, M., (2017 a), Demonstration of translation and scaling procedures: the original data in (a) are centered in (b) and autoscaled in (c). Notice that the autoscaling decreases the between-groups distance in the direction of greatest within-groups scatter and increases it in perpendicular direction in the sense of sphericization of groups. [Figura].
15. Otto, M., (2017 b), 5.15. [Figura].

11.0 ANEXOS

Anexo 1.- Codificación asignada a las muestras usadas.

Muestras Utilizadas en este Trabajo						
Destilados	Licores	T. Blancos	T. Jóvenes	T. Reposados	T. Añejos	T. Extra-añejos
DA 02	LA 03	TB 26	TJ 65	TR 29	TA 13	TEA 27
DA 04	LA 07	TB 43	TJ 2371	TR 65	TA 21	TEA 2429
DA 05	LA 11	TB 113	TJ 2550	TR 139	TA 30	TEA 2483
DA 06	LA 13	TB 115	TJ 2656	TR 2473	TA 101	TEA 2570
DA 08	LA 14	TB 152	TJM 30	TR 2522	TA 2394	TEA 2847
DA 09	LA 17	TB 2968	TJM 2576	TR 2657	TA 2431	TEA 2912
DA 10	LA 18	TB 4743	TJM 2829	TR 3086	TA 2472	TEA 3043
DA 12	LA 19	TB 5405	TJM 2839	TR 4543	TA 2487	TEA 4239
DA 15	LA 23	TB 5588	TJM 4634	TR 4738	TA 2501	TEA 4803
DA 16	LA 47	TB 6907	TJM 4764	TR 4765	TA 2787	TEA 4966
DA 20		TB 6910	TJM 5400	TR 4997	TA 2808	TEA 5426
DA 21		TB 6917	TJM 5448	TR 5402	TA 2815	TEA 5569
DA 22		TB 8539	TJM 5466	TR 5454	TA 2848	TEA 8643
DA 24		TB 8609	TJM 6870	TR 5580	TA 2879	TEA 8650
DA 25		TB 8657	TJM 8706	TR 5592	TA 2884	TEA 8705
DA 26		TB 8680		TR 6887	TA 3090	
DA 27		TB 8710		TR 8596	TA 3095	
DA 28		TB 8718		TR 8623	TA 3125	
DA 29		TB 8762		TR 8631	TA 4184	
		TBM 65		TR 8659	TA 4389	
		TBM 4381		TR 8688	TA 5288	
		TBM 5401		TR 8758	TA 5424	
		TBM 5406		TRM 5090	TA 6868	
		TBM 6872		TRM 5419	TA 6878	
		TBM 6940		TRM 5436	TA 6916	