



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS  
QUÍMICAS

CONTRIBUCIÓN AL ANÁLISIS MULTIVARIABLE DE BEBIDAS  
Y ALIMENTOS POR MÉTODOS ESPECTROMÉTRICOS

INFORME DE TRABAJO DE INVESTIGACIÓN

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRO EN CIENCIAS**

PRESENTA

Q. PEDRO ANZUREZ SÁNCHEZ

TUTOR PRINCIPAL:

Dr. Eduardo Rodríguez de San Miguel Guerrero

Facultad de Química

CIUDAD UNIVERSITARIA, CD. MX., DICIEMBRE DE 2021



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.

## Agradecimientos

A la Universidad Nacional Autónoma de México.

A la Facultad de Química.

Al Dr. Eduardo Rodríguez de San Miguel Guerrero por su tiempo, conocimiento, apoyo incondicional y por permitirme formar parte de su grupo de trabajo.

A la beca otorgada por CONACYT, número de becario 857279.

Al apoyo económico provisto al Fondo Sectorial SAGARPA CONACYT2017-6-292836 "Plataforma nacional de Metabolómica basada en espectroscopia de Resonancia Magnética Nuclear (RMN)-OMICS para la solución de problemas nacionales del sector agroalimentario".

A los miembros del jurado asignado por sus comentarios y acertadas correcciones que ayudaron a enriquecer el trabajo Al gran grupo de trabajo que me permitió crecer con ellos; Jennifer García, Lucía Mora, René González, Yashé Rojas.

## Contenido

1.- Introducción.....	2
2.- Vino .....	2
3.- Miel.....	2
4.- Enfoques analíticos.....	3
I.- Espectroscopia IR-ATR.....	4
II.- Espectroscopia <sup>1</sup> H-RMN .....	4
III.- Espectroscopia UV.....	4
5.- Métodos multivariados .....	4
5.1- PCA.....	5
5.2- PLSDA.....	5
6.- Fusión de datos.....	6
6.1 Índices de evaluación del modelo .....	6
7.- Hipótesis.....	7
8.- Objetivo general .....	7
9.- Objetivos particulares.....	7
10.- Procedimiento experimental.....	8
11.- Discusión de resultados .....	8
I.- Método PCA en IR-ATR.....	10
II- Método PCA en <sup>1</sup> H-RMN .....	11
III.- Método PLSDA con IR-ATR.....	11
IV.- Método PLSDA con espectroscopia <sup>1</sup> H-RMN .....	13
V.- Método fusión de datos <i>Low level</i> .....	14
V.- Método fusión de datos <i>middle level</i> .....	16
12.- Vinos .....	17
I.- Método PCA en IR-ATR y UV .....	18
.....	18
II.- Método PLSDA en IR-ATR y UV.....	18
III.- Fusión de datos <i>low level</i> .....	19
IV Fusión de datos <i>middle level</i> .....	19
13.- Conclusión.....	20

## 1.-Introducción

### 1.1- Trazabilidad de los alimentos

La trazabilidad alimentaria es la capacidad para reconstruir el proceso histórico de producción, transformación y distribución de un alimento, con el fin de detectar rápidamente cualquier problema que pudiera afectar a la salud. En particular, la trazabilidad de los productos naturales se ha convertido en un tema importante considerando la gran cantidad de posibles fuentes de contaminantes o adulteración (Fechner, Hidalgo, Ruiz, Gil y Pellerano, 2020). La determinación de la autenticidad de la comida constituye, a su vez, una de las pruebas más cruciales en el aseguramiento de su control y calidad (Arvanitoyannis, Katsota, Soufleros, Kallithraka & 1999).

### 1.2.- Vino

El vino se puede considerar como una muestra compleja compuestos inorgánicos (contenido mineral) y orgánicos (volátiles, principalmente etanol y no volátiles, principalmente azúcares y ácidos) (Šelih, Šala, and Drgan, 2014), así como una variedad de compuestos químicos como péptidos, proteínas, carbohidratos, tioles y compuestos fenólicos.

El origen geográfico y la autenticidad son factores que influyen en la percepción general de los vinos en términos de calidad y precio (Geana, Iordache, Ionete, Marinescu, Ranca y Culea, 2013) (Fechner, Hidalgo, Ruiz, Gil y Pellerano, 2020), ya que algunos países y regiones son conocidos por producir excelentes vinos de alto valor comercial (Šelih, Šala, and Drgan, 2014), siendo esto de gran importancia para los consumidores y los productores (Geana, I., A., R., A., A. and Culea, M. 2013).

La adulteración del vino, principalmente en términos de variedades y regiones de origen geográfico, es una problemática generalizada (Arvanitoyannis, Katsota, Soufleros, Kallithraka & 1999). Su autenticidad se ha investigado exhaustivamente debido a ser un producto fácilmente adulterable debido a su disponibilidad en todo el mundo. Así se requieren controles meticulosos y continuos para mantener su calidad (Arvanitoyannis, Katsota, Soufleros, Kallithraka & 1999). La autenticidad del vino está garantizada por los estrictos lineamientos establecidos por las autoridades nacionales responsables (por ejemplo, el Instituto Nacional de Denominaciones de Origen en Francia) que incluyen evaluación sensorial oficial, análisis químicos y examen del registro que lleva el vino. (Coetzee, Van Jaarsveld, y Vanhaecke, 2014).

En los últimos años, se ha avanzado mucho en la autenticación de alimentos mediante técnicas de toma de huellas dactilares, en particular en términos de determinación de procedencia. El método combina el análisis químico, por cualquiera de una variedad de técnicas analíticas instrumentales (principalmente el análisis de la proporción de isótopos y elementos traza), y el análisis estadístico multivariado de los datos químicos, para obtener identificación y clasificación de productos alimentarios (Coetzee y Vanhaecke, 2005) (Coetzee, Van Jaarsveld, y Vanhaecke, 2014). Actualmente existe una gran variedad de técnicas combinadas que emplean la clasificación en grupos para identificar la autenticidad del vino (Coetzee, Van Jaarsveld, y Vanhaecke, 2014).

### 1.3.- Miel

La miel es un alimento natural y saludable producido por las abejas (del néctar floral) (Ghanavati, Javaheran, Marini, Nescatelli & Biancolillo 2020), es un importante producto apícola y natural que se caracteriza por un valor comercial creciente (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020). Todas las propiedades de la miel, los valores nutricionales, los efectos beneficiosos y las propiedades organolépticas derivan de su particular composición química (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020). La miel se caracteriza por un valioso contenido de enzimas,

aminoácidos, vitaminas, ácidos orgánicos, minerales antioxidantes, y compuestos fenólicos (Ballabio, & Todeschini, 2009). En busca de néctar las abejas recorren varias plantas, proporcionando miel multifloral, sin embargo, en ocasiones es posible obtener mieles uniflorales puras, las cuales son más raras y valoradas, ya que presentan propiedades sensoriales, melisopalinológicas y fisicoquímicas típicas (Ghanavati, Javaheran, Marini, Nescatelli & Biancolillo 2020). Para asegurar la calidad, el valor comercial (el cual está influenciado por su origen botánico y geográfico) y proteger a los consumidores de la adulteración del producto (Ballabio, & Todeschini, 2009), deben declararse los orígenes tanto botánicos como geográficos del producto (Ghanavati, Javaheran, Marini, Nescatelli & Biancolillo 2020).

La composición química y las características organolépticas de la miel están fuertemente influenciadas por su origen geográfico. Las características de la miel dependen principalmente del origen botánico del néctar. La fuente floral del néctar afecta predominantemente la composición química de la miel en términos de su contenido de proteínas, carbohidratos, enzimas, minerales y ácidos orgánicos. Según la Norma del Codex Alimentarius para la Miel y la Directiva del Consejo de la Unión Europea relacionada con la miel; “Se permite el uso de una denominación botánica de miel si se origina predominantemente de la fuente floral indicada”. En particular, la miel unifloral tiene una mayor demanda y valor comercial en el mercado, ya que son limitados en su producción y requieren de mucho tiempo (Ulloa, Guerra, Cavaco, Da Costa, Figueira, & Brigas, 2013) (Gok, Severcan, Goormaghtingh, Kandemir, & Severca 2015). Desde la perspectiva económica, la autenticación de la miel es necesaria para evitar la competencia desleal que puede conducir a una desestabilización en el mercado.

En el mercado existe una gran variedad de métodos de adulteración como la adición de jarabe de maíz, jarabe de arroz y jarabe de maíz con alto contenido de fructosa. (Li, Zeng, Lin, Zhang, Zhu, Yao, Wang, Yao & Wu, 2020), (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020). La adulteración de la miel se detecta generalmente mediante análisis de proporción de isótopos de carbono estable, resonancia magnética nuclear, cromatografía o análisis térmico diferencial. En los cuatro métodos, el análisis de la relación de isótopos de carbono estable es la técnica analítica estándar para la identificación de miel adulterada con jarabe (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020). El análisis melisopalinológico, es decir, la identificación y cuantificación de los granos de polen en el sedimento de la miel, es el método de referencia para determinar el origen de la miel, sin embargo, este método implica un laborioso procedimiento de recuento que requiere conocimientos especializados y experiencia en la interpretación de los resultados, lo que hace que la identificación del origen botánico de las mieles sea difícil y requiera de mucho tiempo (Ulloa, Guerra, Cavaco, Da Costa, Figueira, & Brigas, 2013). Los métodos de detección anteriores consumen tiempo para su análisis y son destructivos para las muestras, requieren instrumentos costosos que solo pueden ser operados por técnicos profesionales, lo que dificulta la detección rápida en el sitio (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020).

#### 1.4.- Enfoques analíticos

Los estudios de autenticación de muestras de alimentos se pueden realizar utilizando diferentes enfoques analíticos como perfiles de compuestos orgánicos volátiles mediante cromatografía de gases-espectrometría de masas (GC-MS), espectrometría de masas de reacción de transferencia de protones (PTR- MS), Análisis isotópico estable, enfoques espectroscópicos como Infrarrojo con transformada de Fourier - Reflexión Total Atenuada (FTIR- ATR), Infrarrojo (IR), Cercano (NIR), Raman y Resonancia Magnética Nuclear (NMR).

Sin embargo, existe la necesidad de métodos de detección que sean no invasivos, y sí más rápidos, más manejables y adecuados para el cribado de muestras en gran escala (Ulloa, Guerra, Cavaco,

Da Costa, Figueira, & Brigas, 2013). En los últimos años, debido a su naturaleza rápida, sencilla y no destructiva, la espectroscopia se ha utilizado cada vez más en la detección de la adulteración de alimentos. Se han aplicado técnicas de espectroscopia de infrarrojo cercano y medio al estudio de la detección de adulteraciones en alimentos.

#### 1.4.1.- Espectroscopia FTIR-ATR

La espectroscopia FTIR- ATR se ha preferido como técnica rápida, no destructiva, sin reactivos, independiente del operador y barata para la industria alimentaria para la cuantificación de diversas muestras de alimentos (Gok, Severcan, Goormaghtingh, Kandemir, & Severca 2015). La técnica de ATR (Reflexión Total Atenuada) elimina la preparación de la muestra y ofrece un muestreo fácil y rápido de líquidos y sólidos. Esta técnica no destructiva requiere solo una pequeña cantidad de muestras para su análisis. La espectroscopia tiene la ventaja de que puede medir múltiples parámetros simultáneamente en una medición.

El espectro de un material orgánico puede dar una firma global de composición basada básicamente en la evaluación de las estructuras química orgánicas que contienen enlaces O-H, N-H y C-H que, con la aplicación de técnicas quimiométricas se puede utilizar para dilucidar características de composición particulares en la matriz alimentaria que no se detectan fácilmente mediante análisis de composición dirigida.

#### 1.4.2.- Espectroscopia <sup>1</sup>H-RMN

Los requisitos de confiabilidad, rapidez, precisión, consistencia y simplicidad de alimentos para la evaluación de la calidad e inocuidad de los productos alimenticios alentaron el desarrollo de los consumidores de obtener calidades superiores de alimentos. La espectroscopia de <sup>1</sup>H-RMN se puede utilizar en el análisis de alimentos para la discriminación de origen y el descubrimiento de biomarcadores. (Rolin, Deborde, Maucourt, Cabasson, Fauvelle, Jacob, Canlet, & Moing, 2013)

Las señales de RMN tienen un gran número de propiedades medibles. La intensidad, frecuencia, forma de línea, ancho de línea y tiempo de relajación de las señales de RMN son las principales características de interés. La RMN se puede aplicar a muestras y extractos complejos de alimentos y plantas, mediante la cual se pueden identificar y cuantificar los metabolitos (Rolin, Deborde, Maucourt, Cabasson, Fauvelle, Jacob, Canlet, & Moing, 2013).

#### 1.4.3.- Espectroscopia UV

La espectroscopia UV-vis es una técnica basada en medir la absorción de radiación UV y visible por moléculas, siendo la región UV-vis del espectro la longitud de onda que va de 190nm a 800n. La posición espectral de una banda de absorción es indicativa de la presencia o ausencia de ciertas características estructurales o grupos funcionales. Los principales motivos de interés de esta metodología son su amplia aplicabilidad, rapidez de análisis, ausencia de residuos generados y su facilidad de uso, sin costos y sin operador calificado. Tiene una alta sensibilidad, una selectividad moderada-alta y una buena precisión. Todas estas características hacen de la UV-vis una técnica adecuada para el control de los procesos productivos, así como para el seguimiento y evaluación de la composición de la calidad de los productos en muestras de alimentos y bebidas, farmacéuticas y biológicas (Ríos, Azcarate, Camiña, Callejón, & Amigo, 2019).

#### 1.5.- Métodos multivariados

La caracterización físico-química mediante los enfoques analíticos puede servir para construir modelos de clasificación estadística capaces de discriminar las diferentes variedades o áreas geográficas de producción; estos modelos suelen basarse en enfoques multivariados y pueden

aplicarse a muestras desconocidas para predecir su origen geográfico o botánico y así identificar posibles adulteraciones y fraudes.

Los métodos multivariados se han aplicado recientemente en el campo de la autenticación de la miel en diferentes grados, a partir del análisis de componentes principales (PCA), PCA multivariante y métodos de agrupación. También se han aplicado modelos de clasificación multivariantes supervisados, siendo las herramientas más explotadas: Mínimos cuadrados parciales- Análisis discriminantes (PLS-DA), Orthogonal -PLS, Análisis discriminante lineal.

Existen dos enfoques en estadística multivariante para los métodos de reconocimiento de patrones:

- No supervisados: generalmente representa el primer paso en el análisis de datos, se realizan sin conocimiento previo de la muestra para resumir, explorar y descubrir grupos, ayudando a visualizar los datos y encontrar posibles valores atípicos. El método más representativo es el Análisis de Componentes Principales (PCA)
- Supervisados: utilizan el conocimiento previo de las muestras para generar modelos que se centren en los efectos de interés, con el propósito de encontrar uno que pueda hacer predicciones en la clasificación de muestras futuras. El método más común es el Análisis Discriminante de Mínimos Cuadrados Parciales (PLS-DA)

#### 1.5.1- PCA

El PCA es un método de reducción de datos en el que cada espectro, que consta de cientos de valores de absorbancia, se representa mediante un punto en un espacio multidimensional mediante una transformación lineal. Las coordenadas del punto son los componentes principales (PC) y la gráfica obtenida se denomina gráfica de puntuaciones. Cada espectro se puede representar como la suma de varios vectores propios ortonormales ponderados, donde los pesos son los PC. (Gok, Severcan, Goormaghtingh, Kandemir, & Severca 2015), (Maione, Barbosa, & Barbosa 2019). PCA utiliza el aprendizaje no supervisado principalmente para la visualización de datos o el procesamiento de datos antes de que se utilicen técnicas supervisadas. El PCA se basa en la varianza de los datos que se dividen en un pequeño número de componentes principales (PC) mediante transformaciones matemáticas. El primer PC describe la información máxima de los datos y cada PC sucesivo describe la cantidad máxima de la varianza residual. Cada PC es una combinación ortogonal de las variables originales de manera que cubre el máximo de la varianza no contabilizada por los componentes anteriores. (Fechner, Hidalgo, Ruiz, Gil y Pellerano, 2020).

#### 1.5.2- PLSDA

El análisis discriminante de mínimos cuadrados parciales es un método de clasificación discriminante basado en el algoritmo de regresión PLS (Ulloa, Guerra, Cavaco, Da Costa, Figueira, & Brigas, 2013), (Ghanavati, Javaheran, Marini, Nescatelli & Biancolillo 2020), y permite transformar un problema de clasificación en un problema de regresión mediante una matriz de respuesta binaria (Y-Dummy) que codifica la clase- información. A continuación, se realiza la clasificación con base en las respuestas predichas. (Ghanavati, Javaheran, Marini, Nescatelli & Biancolillo 2020). Las variables latentes óptimas (VL) se obtienen de la matriz X mediante la covarianza máxima con la clase de objetivo definida en la matriz Y (Ulloa, Guerra, Cavaco, Da Costa, Figueira, & Brigas, 2013). Tiene como objetivo maximizar la covarianza y debe usarse con precaución porque es una técnica supervisada y siempre conduce a una separación de las diferentes clases, incluso si no hay una diferencia real entre ellas.

En la etapa de modelado, el conjunto de calibración de la matriz X y el vector de clase Y se explota para desarrollar el modelado PLS-DA con los LV óptimos. En el proceso de predicción, el resultado

de la predicción  $Y_n$  se obtiene la ecuación 1, donde B son los coeficientes de regresión y  $X_n$  es la matriz de muestras de validación. (Ulloa, Guerra, Cavaco, Da Costa, Figueira, & Brigas, 2013).

$$Y_n = X_n B \quad (1)$$

Sin embargo, los valores predichos del conjunto de validación no son exactamente 0 ó 1. Por lo tanto, es indispensable conservar el número adquirido con información de categoría. (Ulloa, Guerra, Cavaco, Da Costa, Figueira, & Brigas, 2013)

#### 1.6.- Fusión de datos

La fusión de datos es un método novedoso que integra múltiples fuentes de información para generar un conjunto de datos de muestra más específico y completo. La fusión de datos detecta, correlaciona y combina automáticamente los datos obtenidos de varios sensores. Es capaz de ampliar la información de la muestra y mejorar la precisión del modelo, y se ha utilizado ampliamente en áreas que incluyen el reconocimiento de patrones (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020) (Fechner, Hidalgo, Ruiz, Gil y Pellerano, 2020).

Se pueden fusionar diferentes enfoques analíticos para realizar un perfil químico integral. Se proponen diferentes estrategias de fusión de datos y se clasifican según el nivel en el que produzca la fusión de las distintas fuentes de información:

- Estrategias de alto nivel, la clasificación de las muestras se realiza de forma independiente sobre cada fuente de información; las predicciones proporcionadas por los modelos calibrados en las fuentes individuales se combinan luego.
- En las estrategias intermedias, las características generales se extraen de forma independiente de cada fuente analítica de información y luego estas características se combinan y se utilizan para calibrar un modelo de clasificación y predecir las muestras  
La fusión de datos de nivel medio intermedio se realiza a nivel de característica. Se extrae una característica global de cada fuente de información de forma independiente para construir una matriz de característica  $T_w$  para esa fuente de información. Esta matriz tiene una dimensión de  $(n \times p_w')$ , donde n es el número de muestras y  $p_w'$  es la característica extraída independientemente de la fuente de información w. Todas las matrices de características  $T_w$  se combinan en una matriz global T, que se utiliza como los datos de entrada finales del modelo de clasificación. En comparación con la fusión de datos de bajo nivel, la fusión de datos de nivel intermedio garantiza una representación más equilibrada de cada fuente de información en la matriz T final cuando es necesario fusionar fuentes de información que contienen una gran cantidad de variables. (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020). Ballabio, & Todeschini, 2009)
- La fusión de datos de bajo nivel se realiza a nivel de medición. Las fuentes de información se fusionan en forma de muestras para proporcionar una matriz global X, en la que las filas (n) representan el número de muestras y las columnas (p) representan el número de variables ( $p = \sum_{w=1}^W p_w$ ). Esta matriz X se utiliza como entrada final (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020) (Ballabio, & Todeschini, 2009)  
En los métodos de bajo nivel, las fuentes de información se concatenan directamente y las características generales se extraen del perfil químico fusionado para proporcionar la clasificación de la muestra.

##### 1.6.1 Índices de evaluación del modelo

La evaluación del modelo mide la efectividad de diferentes modelos en el espacio de parámetros y permite la extracción de características. El desempeño de los modelos de clasificación generalmente

se evalúa en términos de precisión, sensibilidad y especificidad del conjunto de predicciones. Los valores cercanos a 1 para estos índices indican un desempeño de clasificación superior al modelo.

$$\text{Precisión} = \frac{n_{\text{correcto}}}{n_{\text{total}}}$$

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

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

$n_{\text{correcto}}$  representa el número de muestras en el conjunto de predicciones clasificadas correctamente por el modelo;  $n_{\text{total}}$  es el número de todas las muestras en el conjunto de predicciones; TP es el número de muestras positivas en el conjunto de predicciones clasificadas correctamente por el modelo; FN es el número de muestras positivas en el conjunto de predicciones clasificadas incorrectamente por el modelo; FP es el número de muestras negativas en el conjunto de predicciones clasificadas incorrectamente por el modelo; y TN es el número de muestras negativas en el conjunto de predicciones clasificadas correctamente por el modelo (Huang, Song, Guo, Guang, Yang, Li, Zhao, y Yang 2020), (Ulloa, Guerra, Cavaco, Da Costa, Figueira, & Brigas, 2013).

La precisión es la relación correcta entre las muestras clasificadas correctamente y las muestras totales. El índice de precisión es uno de los indicadores más importantes para evaluar el desempeño del modelo. Además, también se tienen en cuenta los índices de sensibilidad y especificidad. En realidad, la sensibilidad explica la probabilidad de clasificación corregida de muestras positivas. La especificidad se puede interpretar como el porcentaje de muestras negativas que están correctamente categorizadas.

Los métodos de fusión de datos suelen aprovechar el PCA como método de extracción de características, mientras que la clasificación se suele realizar mediante PLS-DA.

## 2.- Hipótesis

Es posible diferenciar y clasificar muestras de miel y vino mediante fusión de datos, utilizando como fuentes de información de métodos analíticos como  $^1\text{H}$ -RMN, IR-ATR y UV.

## 3.- Objetivo

### 3.1 Objetivo general

Integrar los datos de enfoques analíticos (como IR-ATR,  $^1\text{H}$ -RMN y UV) para construir modelos precisos y robustos para la clasificación de

### 3.2.- Objetivos particulares

- Determinar y evaluar la clasificación de las muestras mediante PCA, así como el mejor preprocesamiento de los datos. Determinar las longitudes de onda que mayor covarianza proporcionen.

- Determinar y evaluar la clasificación de las muestras mediante un método supervisado (PLS-DA), así como compararlo con un método no supervisado (PCA).

- Determinar y evaluar la clasificación de las muestras mediante los métodos de fusión de datos del tipo *low level* y *middle level*. Comparar los métodos para determinar el mejor método de clasificación.

#### 4.- Procedimiento experimental

Se analizaron 17 muestras diferentes de miel unifloral; siete muestras corresponden a miel de mesquite, 6 muestras corresponden a miel de arañjo y cuatro a miel de mantequilla. Las técnicas utilizadas para muestras fueron IR-ATR y <sup>1</sup>H-RMN.

Se analizaron 79 muestras de vino correspondientes a los tipos de vino; La Cetto, Monte Xanic y Santo Tomas. Muestras son analizadas mediante la técnica IR ATR y UV-vis.

Se determinó el mejor modelo que permita clasificar cada tipo de alimento (miel y vino). Las herramientas utilizadas fueron: preprocesamiento de muestra, análisis multivariable supervisados y no supervisados (PLS-DA y PCA) y fusión de datos (nivel bajo y nivel intermedio).

Se utilizaron enfoques de fusión de datos de nivel bajo y medio, utilizando PLSDA como medio de extracción de características, para posteriormente realizar PLS-DA para la clasificación de las mieles uniflorales y vinos. La adopción de una única técnica espectroscópica en la identificación de la miel puede conducir a una precisión y robustez poco deseables en el modelo de identificación de la miel establecido debido a la fuente de información de datos.

Para todos los modelos de PLSDA, las variables se preprocesaron con *mean data* y la primera derivada de *Salgov*, se usó la validación cruzada predeterminada en el paquete PLS\_ToolBox para MATLAB. Todos los modelos supervisados y no supervisados se calcularon con el paquete PLS\_ToolBox para MATLAB.

#### 5.- Discusión de resultados

Se obtuvieron los espectros ATR-IR y de <sup>1</sup>H-RMN de miel de mesquite, naranjo y mantequilla. Los segundos fueron directamente proporcionados por la Dra. Nuria Esturau Escofet del Instituto de Química y constituyen tema de la tesis de licenciatura del alumno Jorge Enrique Sánchez Sánchez. Para ATR-IR se empleó la región de 650 cm<sup>-1</sup> a 4000 cm<sup>-1</sup> con 32 exploraciones y una resolución de 4 cm<sup>-1</sup> y para <sup>1</sup>H-RMN a 700 MHz la región de 9 a 0 ppm. Se contó con diecisiete muestras de miel correspondientes a los 3 tipos diferentes de miel; siete muestras de miel de mezquite, seis muestras de miel naranjo y cuatro muestras de miel de mantequilla.

Se analizaron un total de 17 muestras de mieles mexicanas de diferentes orígenes botánicos y geográficos (Tabla 1) proporcionadas por la ASGANAREN. Las mieles fueron obtenidas directamente de diferentes apiarios y el origen floral de cada una fue declarado por el apicultor. Las muestras de miel se almacenaron en obscuridad en frascos de plástico con tapa de rosca a 20°C hasta su análisis.

Tabla 1.- Información de mieles analizadas

Muestra	Ciudad	Estado	Tipo de miel
01-007-03-01	Rincón de Romos	Aguascalientes	Mezquite
01-007-03-02	Rincón de Romos	Aguascalientes	Mezquite
01-007-03-03	Rincón de Romos	Aguascalientes	Mezquite

01-007-03-04	Rincón de Romos	Agascalientes	Mezquite
01-007-03-05	Rincón de Romos	Agascalientes	Mezquite
01-011-02-02	San Fco. de los Romo	Agascalientes	Mezquite
19-124-28-03	Zapotlanejo	Jalisco	Mezquite
29-013-01-01	Huamantla	Tlaxcala	Mantequilla
29-013-01-19	Huamantla	Tlaxcala	Mantequilla
29-013-01-20	Huamantla	Tlaxcala	Mantequilla
29-013-01-21	Huamantla	Tlaxcala	Mantequilla
30-102-01-03	Mtz. De la Torre	Veracruz	Naranja
30-102-01-06	Mtz. De la Torre	Veracruz	Naranja
30-102-01-07	Mtz. De la Torre	Veracruz	Naranja
30-102-01-08	Mtz. De la Torre	Veracruz	Naranja
30-102-01-17	Mtz. De la Torre	Veracruz	Naranja
30-102-01-18	Mtz. De la Torre	Veracruz	Naranja

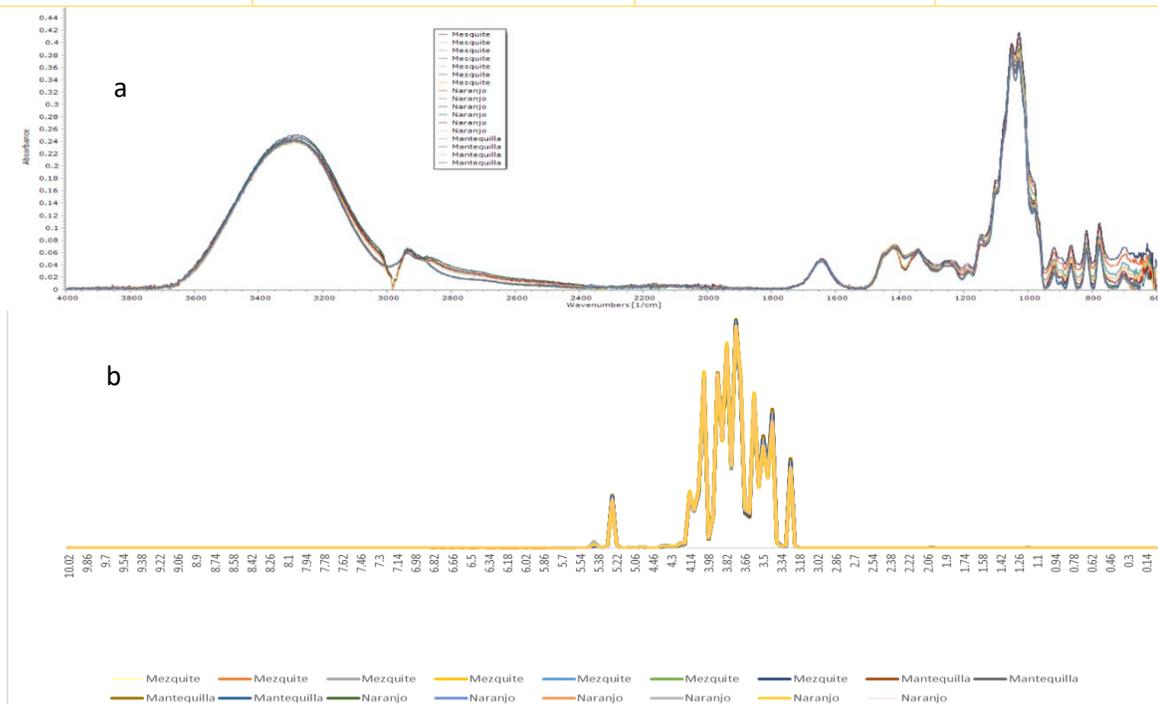


Figura 1.- a) Espectro ATR-IR de diecisiete muestras de miel correspondientes a los 3 tres diferentes tipos de miel; siete muestras de miel de mezquite, seis muestras de miel naranjo y cuatro muestras de miel de mantequilla. b) Espectro <sup>1</sup>H-RMN (700 MHz, H<sub>2</sub>O/D<sub>2</sub>O, 300K) de diecisiete muestras de miel correspondientes a los 3 tres diferentes tipos de miel.

Los espectros ATR-IR de mieles uniflorales (miel de mezquite, naranjo y mantequilla) se muestran en la figura 1 inciso a, en la región de  $650\text{ cm}^{-1}$  a  $4000\text{ cm}^{-1}$ . Se observa en la gráfica inciso a, que las regiones entre  $650$  a  $1300\text{ cm}^{-1}$  y  $2900$  a  $3100\text{ cm}^{-1}$  hay una gran variabilidad entre las señales, las cuales pueden definir la huella dactilar de cada tipo de miel. Li, Zeng, Lin, Zhang, Zhu, Yao, Wang, Yao & Wu (2020), menciona que la región de huella dactilar para mieles puras de acacia, tilo y miel de azufaifo corresponden en las regiones del espectro de  $650$  a  $1300\text{ cm}^{-1}$ . Las señales de  $1030\text{ cm}^{-1}$  corresponden al estiramiento C-O. La banda en la región de  $800\text{ cm}^{-1}$  a  $1500\text{ cm}^{-1}$  se atribuyó principalmente a las absorciones de monosacáridos y disacáridos en la miel. La región entre  $900\text{ cm}^{-1}$  y  $1200\text{ cm}^{-1}$  se atribuyó a la vibración de estiramiento de C-O y C-C de los carbohidratos. Las tres bandas en la región de  $1200\text{ cm}^{-1}$  a  $1470\text{ cm}^{-1}$  se identifican como las vibraciones de flexión de C-C-H, C-O-H y O-C-H. Las dos bandas atribuidas al agua se ubicaron a  $1650\text{ cm}^{-1}$  y  $3300\text{ cm}^{-1}$  respectivamente, ya que corresponden a las vibraciones H-O-H y al estiramiento O-H respectivamente (Gok, Severcan, Goormaghtingh, Kandemir, & Severca 2015).

Los espectros  $^1\text{H}$ -RMN de mieles uniflorales. (miel de mezquite, naranjo y mantequilla) se en la Gráfica 1 inciso b, donde las señales generadas se encuentra de  $3.0$  a  $5.5\text{ ppm}$ .

Se realizó el análisis de componentes principales (PCA) para la reducción de dimensiones a fin de encontrar un sistema de coordenadas más significativo para los resultados de la matriz de datos. Se realizó un análisis exploratorio utilizando PCA para una disposición matricial de los resultados que contenía 17 filas y 3351 columnas para el método analítico d IR-ATR y para RMN 17 filas y 239 columnas.

#### I.- Método PCA en IR-ATR

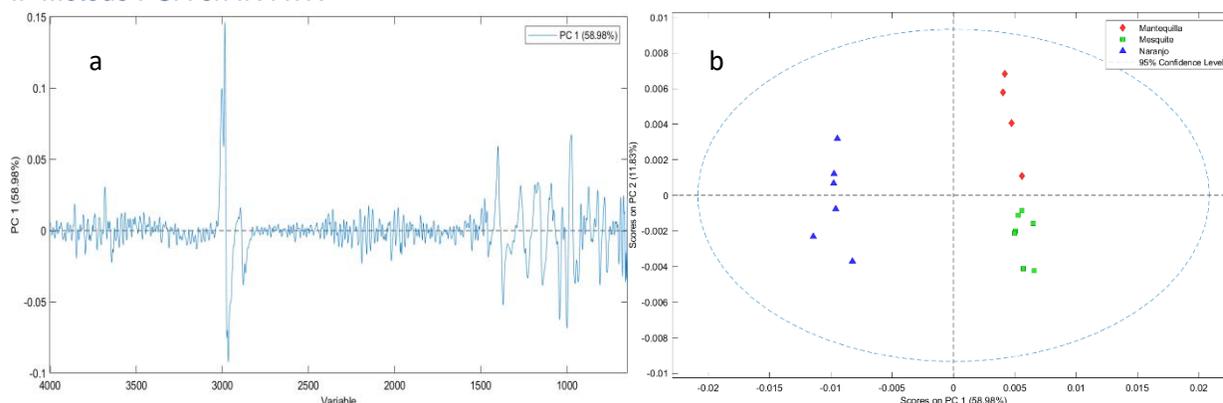


Figura 2.- a) Gráfica de carga del primer componente de PCA para espectroscopia IR-ATR. b) Gráficos de dispersión de PCA para las muestras de miel de mesquite, naranjo y mantequilla sobre  $4000\text{ cm}^{-1}$  a  $650\text{ cm}^{-1}$

En la figura 2a se describe la región espectral de mayor influencia en la clasificación, las cuales corresponde a las regiones que van de  $650\text{ cm}^{-1}$  a  $1500\text{ cm}^{-1}$  y de  $2900$  a  $3100\text{ cm}^{-1}$ . Las grandes cargas negativas y positivas reflejan la contribución de cada variable (Gok, Severcan, Goormaghtingh, Kandemir, & Severca 2015). En la figura 2b se puede visualizar la separación altamente exitosa de 3 conjuntos de datos, los cuales corresponden a las mieles de mesquite, naranjo y mantequilla. El componente PC1 explica el 58.98% de la varianza total del modelo, el componente PC2 describe el 11.83% de la varianza total. El PCA es una herramienta de visualización, la cual se utilizar para identificar variables que se pueden utilizar como patrones y

usarlos datos de entrada para los modelos de análisis discriminantes (Maione, Barbosa, & Barbosa 2019).

## II- Método PCA en <sup>1</sup>H-RMN

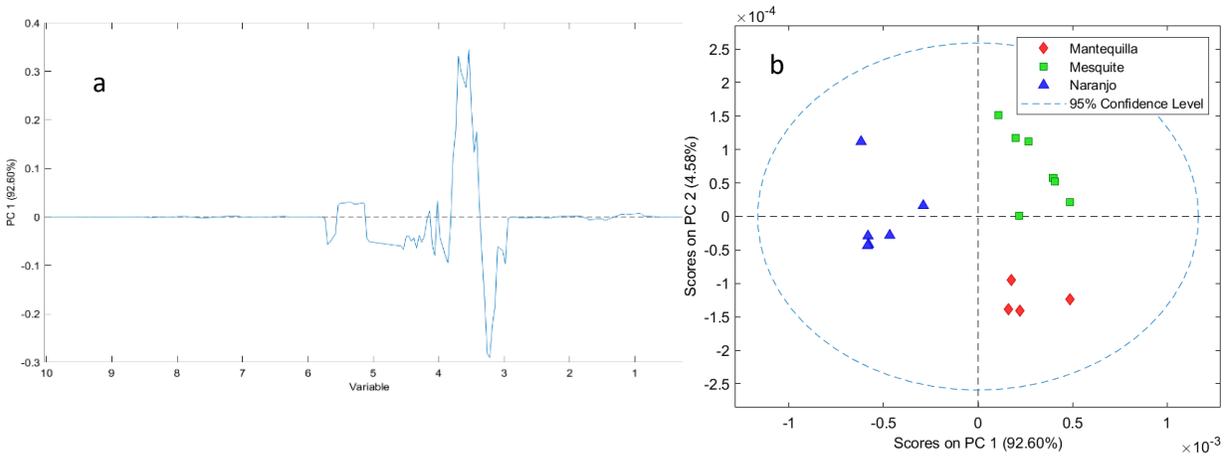


Figura 3.- a) Gráfica de carga del primer componente de PCA para espectroscopia <sup>1</sup>H-RMN b) Gráficos de dispersión de PCA para las muestras de miel de mesquite, naranja y mantequilla.

En la igura 3a se describe la región espectral de mayor influencia en la clasificación, las cuales corresponde entre 2.8 y 5.5 ppm. Las grandes cargas negativas y positivas reflejan la contribución de cada variable (Gok, Severcan, Goormaghtingh, Kandemir, & Severca 2015). En la igura 2b se puede visualizar la separación altamente exitosa de 3 conjuntos de datos, los cuales corresponden a las mieles de mesquite, naranja y mantequilla. El componente PC1 explica el 92.6% de la varianza total del modelo, el componente PC2 describe el 4.58% de la varianza total.

## III.- Método PLSDA con IR-ATR

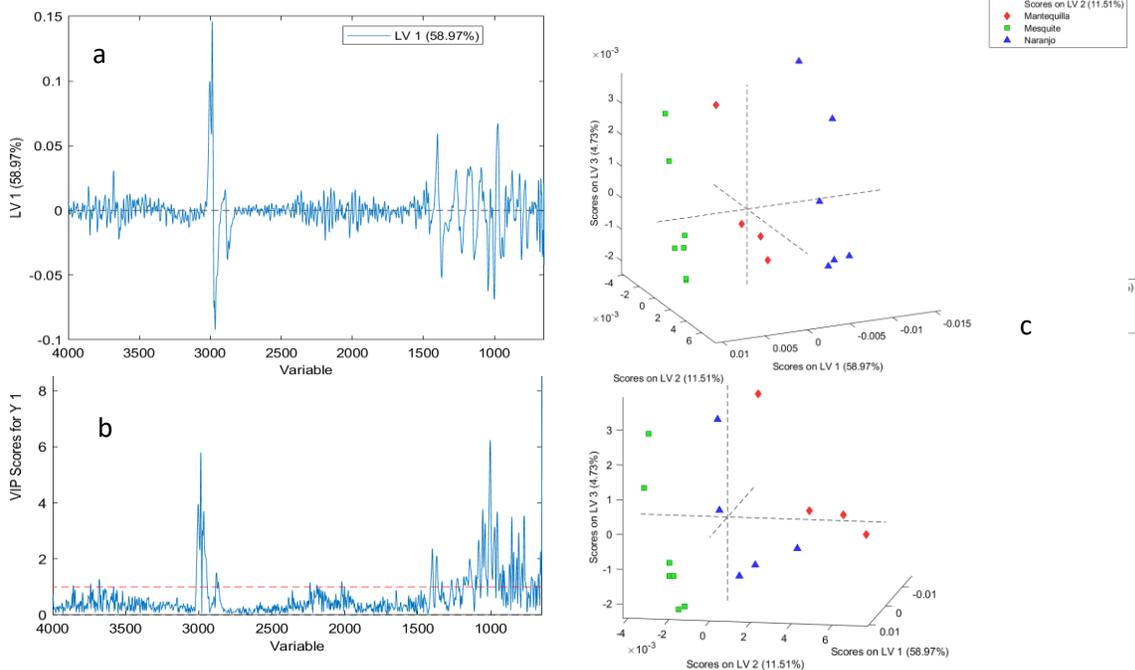


Figura 4.- a) Gráfica de carga de la primera variable latente en espectroscopia IR-ATR b) Análisis VIP c) Gráficos de dispersión de PCA para las muestras de miel de mesquite, naranjo y mantequilla.

El algoritmo PLS-DA descompuso la información de espectros y categorías simultáneamente con el propósito de extraer la información efectiva al máximo. Se eligen tres LV para poder representar de forma tridimensional los gráficos (Li, Zeng, Lin, Zhang, Zhu, Yao, Wang, Yao & Wu, 2020). En la práctica, es más conveniente trazar gráficos 2D o 3D de los LV, dicho gráfico es la proyección de un gráfico multidimensional en un espacio bidimensional o tridimensional. En la figura 4a, se describe la región espectral de mayor influencia en la clasificación, las cuales corresponde a las regiones que van de  $650\text{ cm}^{-1}$  a  $1500\text{ cm}^{-1}$  y de  $2900$  a  $3100\text{ cm}^{-1}$ . La variable latente LV1 explica el 58.97% de la varianza total del modelo, LV2 describe el 11.51% y LV 4.73%. Para Gok, Severcan, Goormaghtingh, Kandemir, & Severca (2015), la separación se logra debido al contenido de agua, el contenido de proteínas y las interacciones agua carbohidratos en cada muestra.

El análisis de Importancia de Variable en Proyección (VIP), se aplicó al modelo PLSDA con el propósito determinar las variables que contribuyen más a la clasificación de cada miel (Ghanavati, Javaheran, Marini, Nescatelli & Biancolillo 2020). Existe una fuerte concordancia entre las variables obtenidas del análisis VIP y las variables obtenidas en la gráfica de carga de VL1. Las variables que atraviesan la línea roja, presentan un índice VIP mayor a 1, dichas variables se encuentran entre las longitudes de onda de  $650\text{ cm}^{-1}$  a  $1300\text{ cm}^{-1}$  y  $2900$  a  $3100\text{ cm}^{-1}$ , por lo tanto contribuyen al modelo de clasificación.

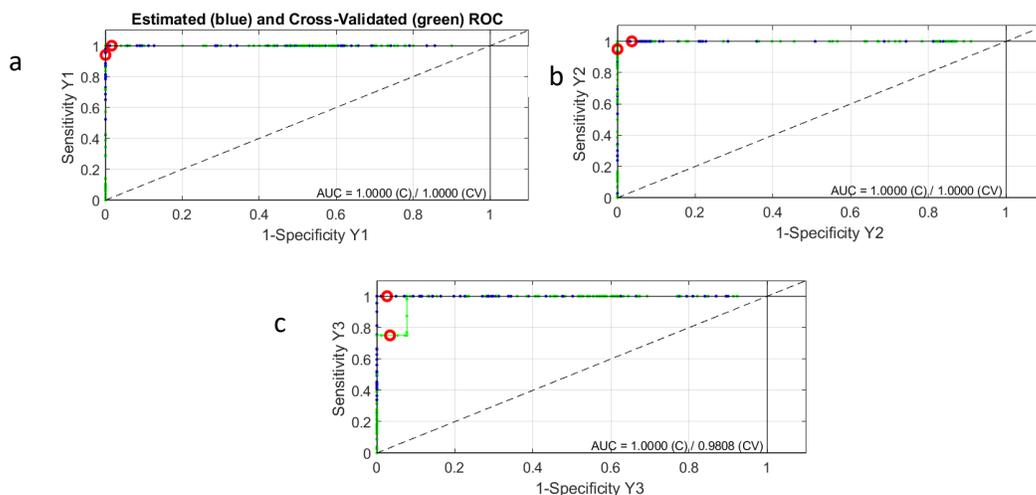


Figura 5.- a) Gráfica ROC correspondiente a la miel de mesquite, b) Gráfica ROC correspondiente a la miel de naranjo c) Gráfica ROC correspondiente a la miel de mantequilla.

Un modelo efectivo no solo está caracterizado por su alta precisión sino también por una alta sensibilidad y especificidad. La sensibilidad y la especificidad son parámetros cruciales para considerar en un modelo de clasificación por PLS-DA. La curva de características operativas del receptor (ROC) es un gráfico de la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (1 menos especificidad) que cuantifica la capacidad discriminatoria de un modelo. El área bajo la curva ROC (AUC) puede usarse como un estimador de la separabilidad de las clases (Li, Zeng, Lin, Zhang, Zhu, Yao, Wang, Yao & Wu, 2020).

-AUC < 0.5 indica que el modelo discriminante es mal clasificador

-AUC > 0.5 Indica que el modelo es un buen clasificador

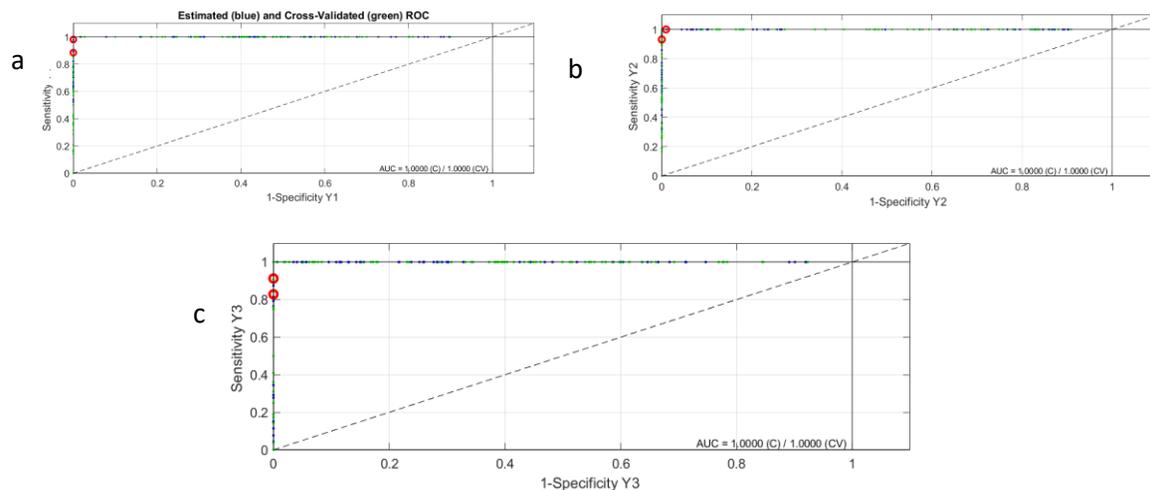
En la figura 5, las gráficas (ROC) a, b y c, la línea azul muestra una sensibilidad de uno, es decir, el porcentaje de muestras positivas que fueron identificadas correctamente (100%). De forma análoga, la tasa de falsos positivos representa la fracción de muestras negativas que se clasifican erróneamente como positivos (0%). Se deduce entonces que el modelo clasifica correctamente las muestras de miel correspondiente a mesquite, naranjo y mantequilla. Para las gráficas a,b y c, la línea verde corresponde a la validación cruzada del modelo.

Miel	Estimado	Validación cruzada (CV)	Calidad del modelo
Mezquite	1	1	Buen clasificador
Naranja	1	1	Buen clasificador
Mantequilla	1	0.9808	Buen clasificador

Tabla 2. Correspondiente a los Valores de AUC para PLSDA-IRATR.

Con base en los datos obtenidos de AUC presentados en la tabla 2, se determinó que el modelo clasifica correctamente las muestras de miel correspondiente a mesquite, naranjo y mantequilla.

#### IV.- Método PLSDA con espectroscopia <sup>1</sup>H-RMN



En la figura 6, las gráficas (ROC) a, b y c, la línea azul muestra una sensibilidad de uno, es decir, el porcentaje de muestras positivas que fueron identificadas correctamente (100%). De forma análoga, la tasa de falsos positivos representó la fracción de muestras negativas que se clasifican erróneamente como positivos (0%). Por lo que el modelo clasifica correctamente las muestras de miel correspondiente a mesquite, naranjo y mantequilla. Para las gráficas a,b y c, la línea verde corresponde a la validación cruzada del modelo.

Miel	Estimado	Validación cruzada (CV)	Calidad del modelo
Mezquite	1	1	Buen clasificador
Naranja	1	1	Buen clasificador
Mantequilla	1	1	Buen clasificador

Tabla 3. Correspondiente a los valores AUC para PLSDA: <sup>1</sup>H-RMN

Con base en los datos obtenidos de AUC presentados en la Tabla 3, se determinó que el modelo clasifica correctamente las muestras de miel correspondiente a mezquite, naranja y mantequilla.

V.- Método fusión de datos *Low level*

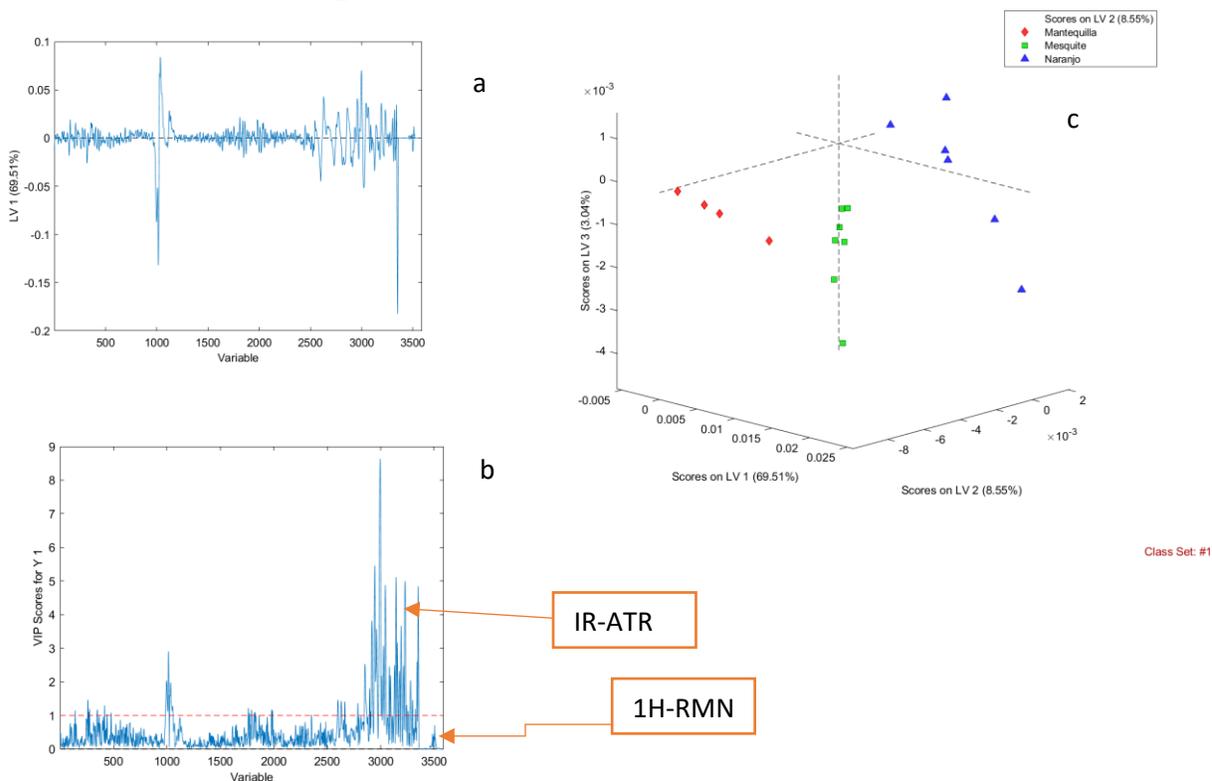


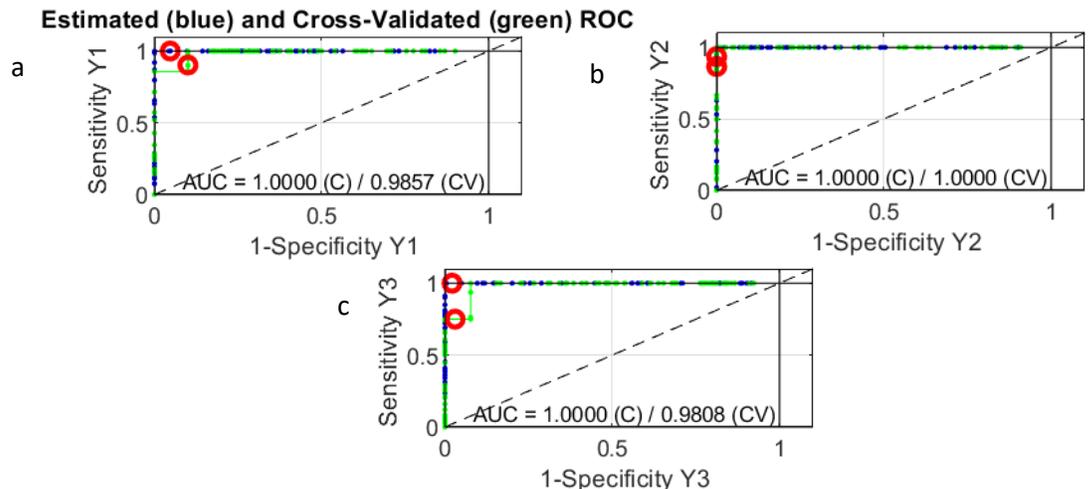
Figura 7.- a) Gráfica de carga de la primera variable latente, con enfoque de fusión de datos *low level* b) Análisis VIP, con enfoque de fusión de datos *low level*. c) con enfoque de Gráficos de dispersión de LV, fusión de datos *low level*.

Al manejar un conjunto de datos de varios bloques, es posible analizar individualmente cada bloque de datos, pero esta no siempre es la solución más adecuada. Hoy en día está bien establecido que es más poderoso manejar datos multiplataforma mediante enfoque de fusión de datos (Ghanavati, Javaheran, Marini, Nescatelli & Biancolillo 2020). Con el propósito de saber si el modelo de

clasificación puede mejorarse, se realizó un modelo de clasificación con un enfoque de fusión de datos *low level*. Los bloques individuales (IR-ATR y <sup>1</sup>H-RMN) se concatenaron para proporcionar una matriz de datos generales con 3590 variables, siendo los diferentes bloques caracterizados por el siguiente número de variables: 3351 para IR y 239 para <sup>1</sup>H-RMN. Los datos de la nueva matriz se encuentran ordenados de la siguiente manera: IR-ATR – <sup>1</sup>H-RMN.

Se eligen tres LV para poder representar de forma tridimensional el modelo matemático. En la figura 6a, se observan dos divisiones, la primera corresponde a IR-ATR y la segunda a <sup>1</sup>H-RMN, la región espectral de mayor influencia en la clasificación corresponde a IR-ATR, las cuales corresponde al eje 1000 y entre 2800 a 3400. La variable latente LV1 explica el 69.51% de la varianza total del modelo, LV2 describe el 8.55% y LV3 3.04%.

Existe una fuerte concordancia entre los variables obtenidas del análisis VIP y las variables obtenidas en la gráfica de carga de LV1. Las variables que atraviesan la línea roja, presentan un índice VIP mayor a 1, dichas variables se encuentran en el 100 y entre 2800 a 3400. Se puede observar que para el enfoque de fusión de datos *low level*, la espectroscopia AT-RMN es quien determina la variabilidad del modelo de clasificación PLSDA.



En la figura 8, las gráficas (ROC) a, b y c, la línea azul muestra una sensibilidad de uno, es decir, el porcentaje de muestras positivas que fueron identificadas correctamente (100%), e forma análoga, la tasa de falsos positivos representó la fracción de muestras negativas que se clasifican erróneamente como positivas (0%). Por lo que el modelo, clasifica correctamente las muestras de miel correspondiente a mesquite, naranjo y mantequilla. Para las gráficas a,b y c, la línea verde corresponde a la validación cruzada del modelo.

Miel	Estimado	Validación cruzada (CV)	Calidad del modelo
Mezquite	1	0.9857	Buen clasificador
Naranjo	1	1	Buen clasificador
Mantequilla	1	0.9808	Buen clasificador

Tabla 4. Correspondiente a los Valores de AUC, correspondiente al método de fusión *low level*.

Con base en los datos obtenidos de AUC presentados en la tabla 3, se determinó que el modelo clasifica correctamente las muestras de miel correspondiente a mesquite, naranjo y mantequilla.

V.- Método fusión de datos *middle level*

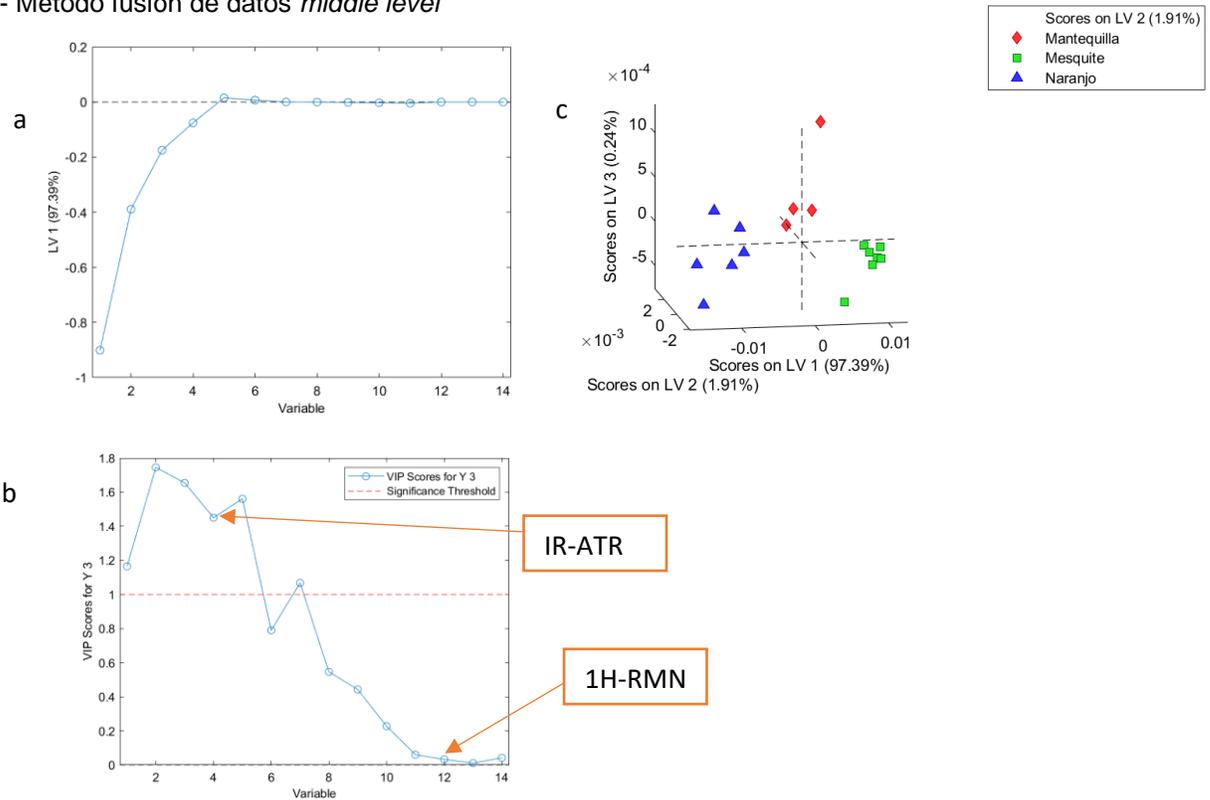


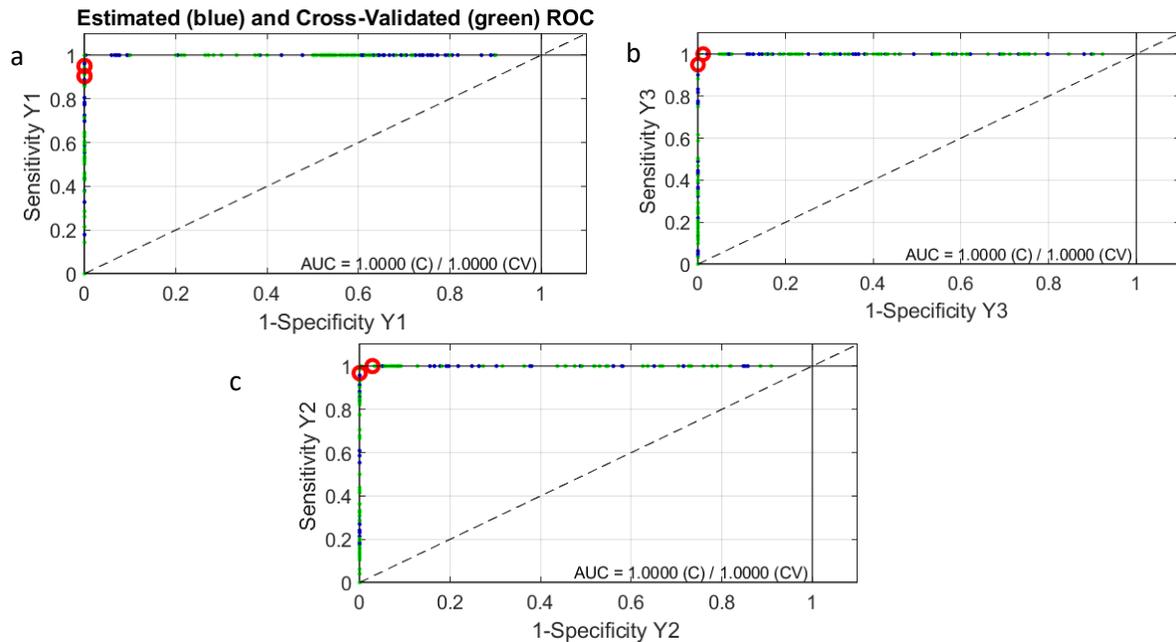
Figura 9.- a) Gráfica de carga de la primera variable latente, con enfoque de fusión de datos *middle level* b) Análisis VIP, con enfoque de fusión de datos *middle level*. c) con enfoque de Gráficos de dispersión de LV, fusión de datos *middle level*.

La extracción de característica para la fusión de datos de nivel intermedio se basó en PLS-DA. Las puntuaciones de las primeras siete LV calculadas en cada bloque individual por separado se concatenaron para proporcionar una matriz que consta de 14 variables (14 LV por 2 bloques). Las variables latentes de la nueva matriz se encuentran ordenados de la siguiente manera: IR-ATR – <sup>1</sup>H-RMN.

En la figura 7a, se observan dos divisiones, la primera corresponde a IR-ATR y la segunda a <sup>1</sup>H-RMN, de la variable una a siete corresponde a las LV del bloque de IR-ATR y las LV de siete a catorce corresponde al bloque de H-RMN. Se observa que las variables latentes que generan mayor variabilidad en el modelo corresponden a la espectroscopia IR-ATR. La variable latente LV1 explica el 97.39% de la varianza total del modelo, LV2 describe el 1.31% y LV3 0.24%.

Existe una fuerte concordancia entre los variables obtenidas del análisis VIP y las variables obtenidas en la gráfica de carga de LV1. Se puede observar que para el enfoque de fusión de datos *middle level*, la espectroscopia AT-RMN es quien determina la variabilidad del modelo de clasificación PLS-DA.

Se debe considerar el uso de RMN ya que es una técnica de gran versatilidad, pero en comparación con la espectroscopia IR-ATR, es más costosa.



En la figura 10, las gráficas (ROC) a, b y c, la línea azul muestra una sensibilidad de uno, es decir, el porcentaje de muestras positivas que fueron identificadas correctamente (100%), De forma análoga, la tasa de falsos positivos representa la fracción de muestras negativas que se clasifican erróneamente como positivos (0%). Por lo que el modelo clasifica correctamente las muestras de miel correspondiente a mesquite, naranjo y mantequilla. Para las gráficas a,b y c, la línea verde corresponde a la validación cruzada del modelo.

Miel	Estimado	Validación cruzada (CV)	Calidad del modelo
<b>Mezquite</b>	1	1	Buen clasificador
<b>Naranjo</b>	1	1	Buen clasificador
<b>Mantequilla</b>	1	1	Buen clasificador

Tabla 5. Correspondiente a los Valores de AUC, correspondiente al método de fusión *middle level*.

Con base en los datos obtenidos de AUC presentados en la tabla 3, se determinó que el modelo clasifica correctamente las muestras de miel correspondiente a mesquite, naranjo y mantequilla.

## 12.- Vinos

## I.- Método PCA en IR-ATR y UV

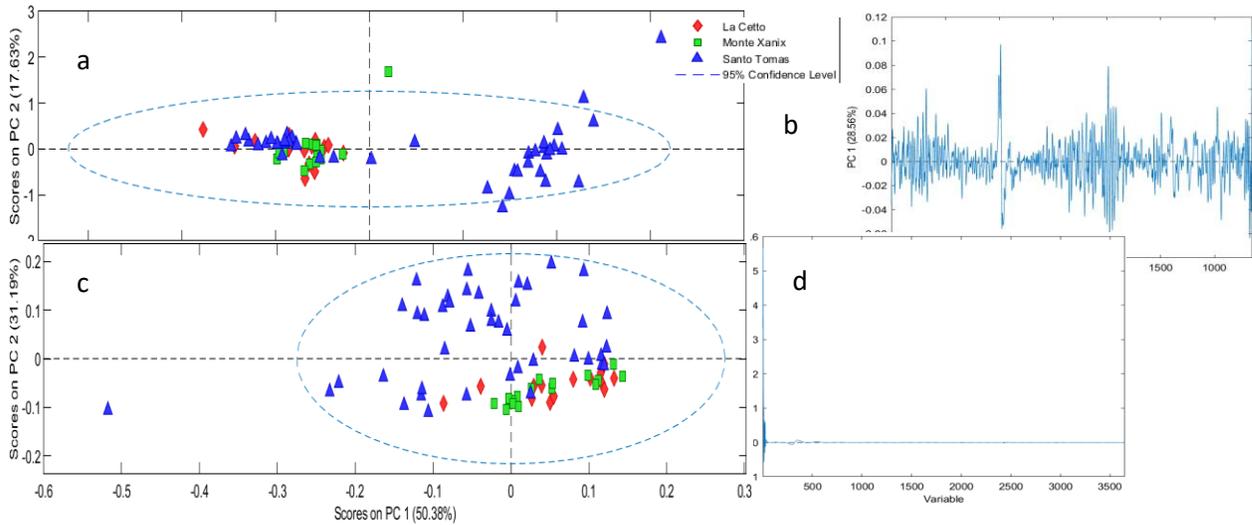
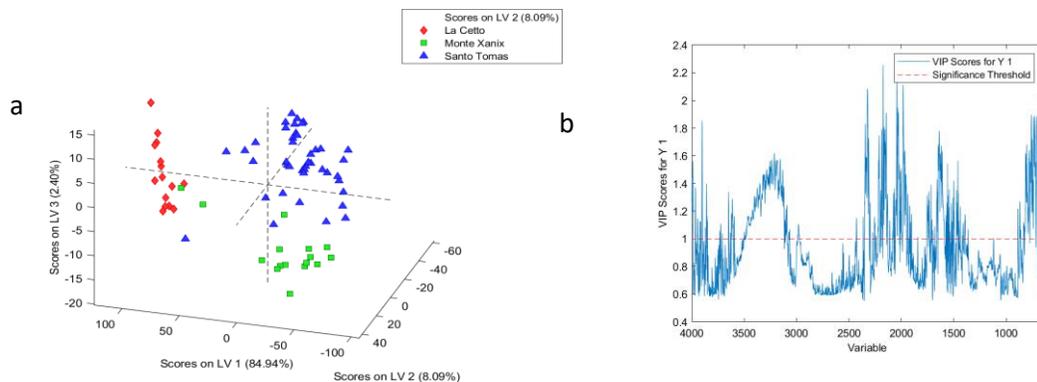


Figura 11.- a) Gráfica de carga del primer componente de PCA para espectroscopia IR-ATR b) Gráfica de carga de PCA para las muestras de vino; La Cetto, Monte Xanic y Santo Tomas para espectroscopia IR-ATR. c) Gráfica de carga del primer componente de PCA para espectroscopia UV b) Gráfica decarga de PCA para las muestras de vino; La Cetto, Monte Xanic y Santo Tomas para espectroscopia UV

Se observa en la figura 11 incisos a y c que existe una nula separación entre las muestras de vino Monte Xanic y La Cetto. Para la espectroscopia IR-ATR el gráfico PCA solo representa el 46.19% de la variabilidad del modelo, el cual es muy bajo en comparación con el correspondiente al gráfico PCA de UV con una variabilidad total de 81.57%. La gráfica de carga b correspondiente al método PCA-IR-ATR, describe la región espectral de mayor influencia en la clasificación, el cual corresponde a todo el espectro de IR-ATR, para inciso d correspondiente al método PCA-UV no muestra significancia del espectro hacia el modelo no supervisado de PCA.

Se procedió entonces a realizar un análisis supervisado para una separación de muestras de vino; La Cetto, Monte Xanic y Santo Tomas.

## II.- Método PLSDA en IR-ATR y UV



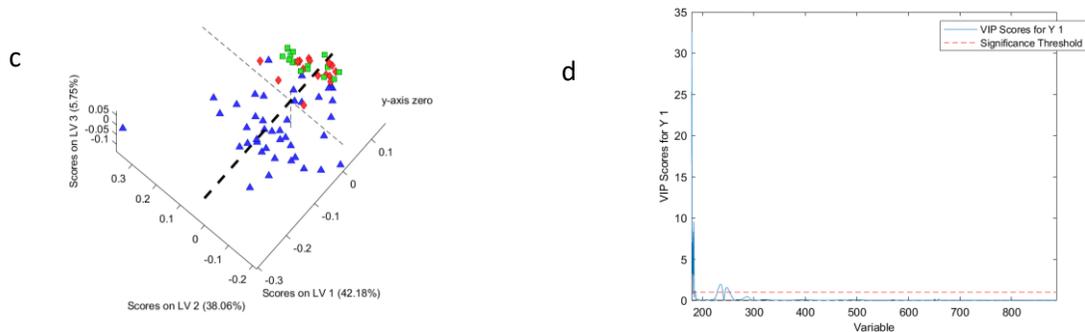


Figura 12.- a) Gráfica de variables latentes de PLSDA para espectroscopia IR-ATR b) Gráfica de VIP de PLSDA para las muestras de vino; La Cetto, Monte Xanic y Santo Tomas para espectroscopia IR-ATR. c) Gráfica de variables latentes de PLSDA para espectroscopia UV b) Gráfica VIP de PLSDA para las muestras de vino; La Cetto, Monte Xanic y Santo Tomas para espectroscopia UV

El inciso a de la Figura 12 muestra una separación entre muestras de vino, generando cuatro grupos; dos para Santo Tomas, uno a La Cetto y otro a Monte Xanic, las variables latentes representan el 95.5 % de la variabilidad del modelo, para el inciso c las variables latentes representan el 85.99%, para dicho modelo no se aprecia una separación entre los tipos de vino La Cetto y Monte Xanic. La gráfica VIP correspondiente al modelo a PLSDA-IRATR muestra la mayor parte del espectro que proporciona variabilidad al modelo supervisado. Para el modelo PLSDA-UV, la gráfica VIP indica que el rango de 230 a 270 nm<sup>-1</sup> es quien proporciona la información al modelo.

### III.- Fusión de datos *low level*

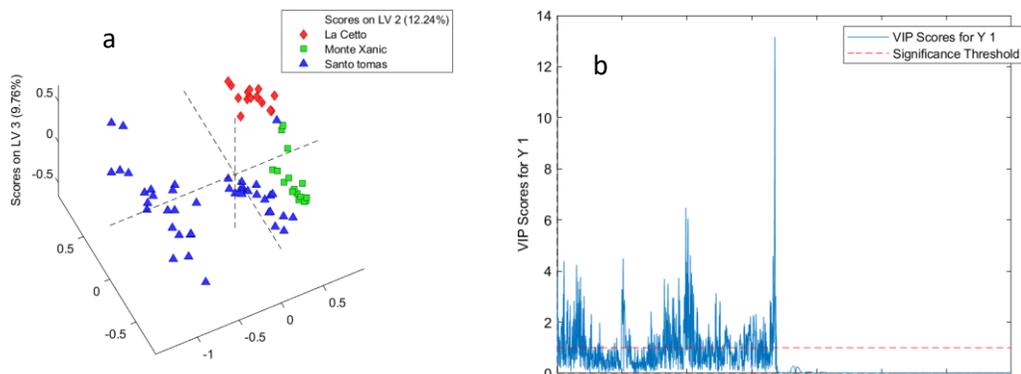


FIGURA 13.- a) Gráfica de variables latentes de PLSDA para un enfoque de fusión de datos b) Gráfica de VIP de PLSDA para un enfoque de fusión; La Cetto, Monte Xanic y Santo Tomas.

Se realiza un enfoque de fusión de datos *low level*, con una matriz concatena de 73 filas y 6999 columnas correspondientes a los enfoques analíticos de IRATR y UV. En la figura 13 a) se observa una correcta clasificación de muestras, sin embargo, es un modelo poco confiable debido a que las variables latentes sólo expresan el 52.32 % de la variabilidad de los datos. El gráfico VIP muestra que la mayor contribución al modelo de PLSDA corresponde al espectro de IRATR y nula para UV.

### IV Fusión de datos *middle level*

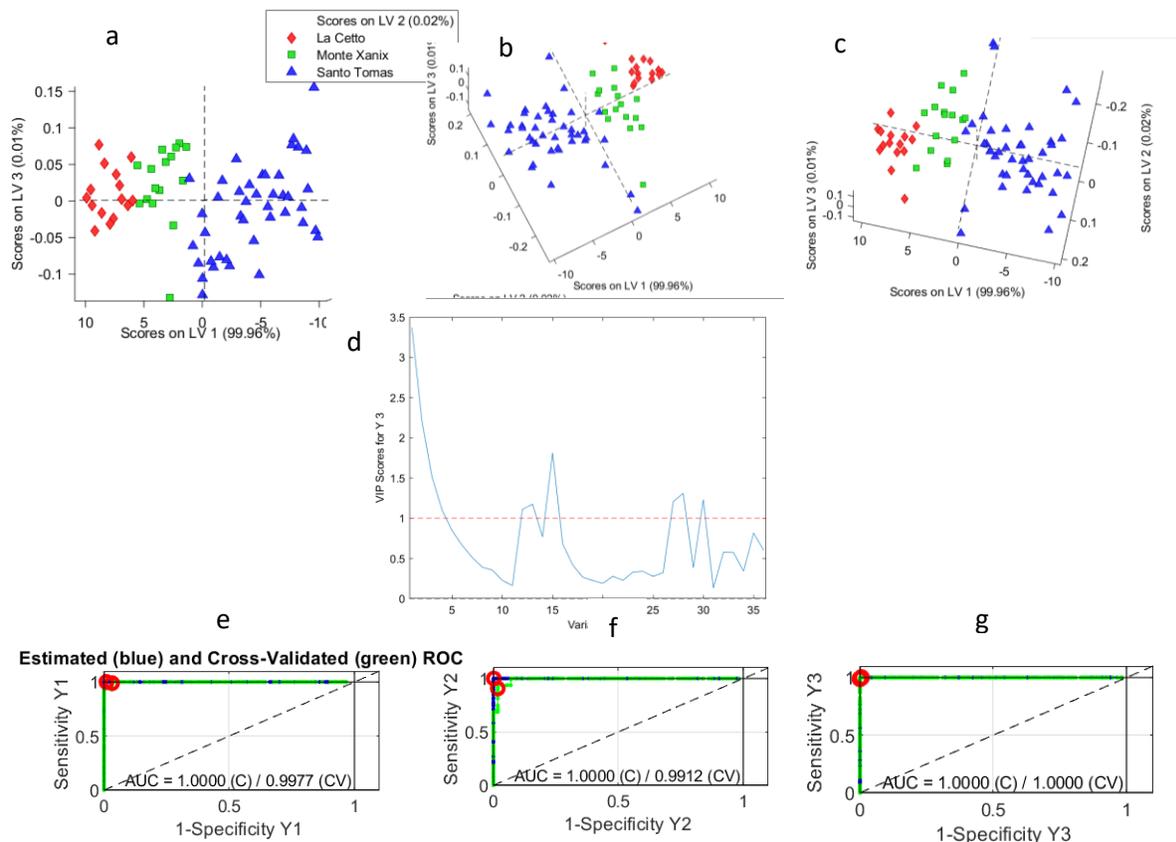


FIGURA 14.- Las a), b) y c) representan las variables latentes de PLS-DA, desde diversas perspectivas, para un enfoque de fusión de datos d) Gráfica de VIP de PLS-DA para un enfoque de fusión middle level; La Cetto, Monte Xanic y Santo Tomas. Incisos e), f) y g) representan las gráficas ROC del enfoque de fusión de datos *middle level*.

Tomando 18 VL de cada modelo se realiza un enfoque de fusión de datos *middle level*, con una matriz concatenada de 73 filas y 36 columnas correspondientes a las variables latentes de los enfoques analíticos de IRATR y UV. Cada uno contribuye con 18 variables latentes. Se observa una clasificación para cada tipo de vino, Santo Tomas forma un solo grupo en comparación con los modelos anteriores que presentaban dos grupos. Hay una clara separación entre los tipos de vino Monte Xanic y La Cetto. La gráfica VIP indica que las VL correspondientes a enfoque analítico IRATR tienen una mayor contribución con el modelo de fusión. Las curvas ROC tienen un valor mayor a 0.5 por lo que se considera que son modelos adecuados para la clasificación de muestras de vino.

### 6.- Conclusiones

A partir de los espectros de <sup>1</sup>H- RMN, IR-ATR y UV fue posible realizar modelado supervisado y no supervisado de muestras de mieles y vinos. las mieles mediante cualquiera de las plataformas analíticas, siendo la espectroscopia infrarroja la que mejor se desempeñó. Se pudo comparar el enfoque de fusión de low level y middle level, donde el enfoque middle level proporcionó la mayor precisión, sensibilidad y especificidad.

Para las muestras de vino, la cantidad de muestras fue mayor por lo es más representativa para un modelo de clasificación. No se pudo observar una separación de grupos mediante PCA; incluso utilizando un modelo supervisado como PLSDA no fue posible la separación de I de vino de Monte Xanic y La Cetto. Las variables latentes mostraron en general un bajo contenido de información para el modelo de clasificación. Se evaluaron los vinos por medios de enfoques de fusión de datos, donde el enfoque *low level* proporcionó una baja capacidad para clasificar las muestras. Se demostró que el enfoque analítico de UV para PLSDA y *low level*- PLSDA no proporciona una buena herramienta clasificatoria.

Sin embargo, mediante el enfoque *middle level*, fue posible clasificar con alta capacidad los tipos de vino. Se demostró que usar variables latentes extraídas de un método supervisado permite clasificar un gran número de muestras de diferentes tipos.

## 7.- Referencias

- Acevedo, F., Jiménez, J., Maldonado, S., Domínguez, E. & Narvaéz, A. (2007). Classification of Wines Produced in Specific Regions by UV-Visible Spectroscopy Combined with Support Vector Machines. *Journal of Agricultural and Food Chemistry*, 55, 6842-6849. <https://doi.org/10.1021/jf070634q>
- Apramita, D., Jitender, J. & Anu-Appaiah, K. (2018). Chemical characterization complemented with chemometrics for the botanical origin identification of unifloral and multifloral honeys from India. *Food Research International*, 107, 216-226. <https://doi.org/10.1016/j.foodres.2018.02.017>
- Arvanitoyannis, I., Katsota, M., Pesarra, E., Soufleros, E. and Kallithraka, S. (1999). Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). *Trends in Food Science & Technology*, 10(10), pp.321-336.
- Bajoub, A., Medina, S., Gómez, M., El, A., Bagur, M., Fernández, A. & Carrasco, A. (2017). Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusión and chemometrics. *Food Chemistry: Molecular Science*, 215, 245-255. <https://doi.org/10.1016/j.foodchem.2016.07.140>
- Ballabio, D. & Todeschini, R. (2009). Cap.4 Classification for Qualitative Analysis. En *Infrared Spectroscopy for Food Quality Analysis and Control* (pp. 83-102). Milan, Italia: Elsevier Inc. DOI: 10.1016 / B978-0-12-374136-3.00004-3
- Ballabio, D., Robotti, E., Grisoni, F., Quasso, F., Bobba, M., Vercelli, S., Gosetti, F., Calabrese, G., Sangiorgi, E., Orlandi, M. & Marengo, E. (2018). Chemical profiling and multivariate data fusion methods for the identification of the botanical origin of honey. *Food Chemistry: Molecular Science*, 266, 79-89. <https://doi.org/10.1016/j.foodchem.2018.05.084>
- Belmonte, J., Romero, R., Martinez, J., Arrebolla, F. & Garrido, A. (2020). <sup>1</sup>H NMR and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical análisis. *Food Chemistry: Molecular Science*, 317. <https://doi.org/10.1016/j.foodchem.2020.126363>.
- Boccard, J. & Rudaz, S. (2020). Cap. 4.2 Analysis of Metabolomics Data-A Chemometrics Perspective. En *Comprehensive Chemometrics* (2ª ed.). (pp. 483-505). Lausanne, Suiza: Elsevier B.V. <https://doi.org/10.1016/B978-0-12-409547-2.14593-7>
- Borràs, E., Ferré, J., Boqué R., Mestres, M., Aceña, L., Calvo, A. & Busto, O. (2016). Olive oil sensory defects classification with data fusion of instrumental techniques and multivariate analysis (PLS-DA). *Food Chemistry: Molecular Science*, 203, 314-322. <https://doi.org/10.1016/j.foodchem.2016.02.038>

- Coetzee, P. & Vanhaecke, F. (2005). Classifying wine according to geographical origin via quadrupole-based ICP–mass spectrometry measurements of boron isotope ratios. *Química analítica y bioanalítica*, 383, 977-984. DOI: 10.1007 / s00216-005-0093-7
- Coetzee, P., Steffens, E., Eiselen, R., Augustyn, O., Balcaen, L. & Vanhaecke, F. (2005). Multi-element Analysis of South African Wines by ICP-MS and Their Classification According to Geographical Origin . *Juornal of Agricultural and Food Chemistry*, 53, 5060-5066. DOI: 10.1021 / jf048268n
- Coetzee, P., van Jaarsveld, F. and Vanhaecke, F. (2014). Intraregional classification of wine via ICP-MS elemental fingerprinting. *Food Chemistry*, 164, 485-492. <https://doi.org/10.1016/j.foodchem.2014.05.027>
- Fechner, D., Hidalgo, M., Ruiz, J., Gil, E & Pellerano, R. (2020). Geographical origin authentication of honey produced in Argentina. *Food Bioscience*, 33, 100483. <https://doi.org/10.1016/j.fbio.2019.100483>
- Ghanavati, S., Javaheran, M., Marini, F., Nescatelli, R. & Biancolillo, A. (2020). Classification of honey applying high performance liquid chromatography, near-infrared spectroscopy and chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 202, 104037. <https://doi.org/10.1016/j.chemolab.2020.104037>
- Gok, S., Severcan, M., Goormaghtingh, E., Kandemir, I. & Severcan. F. (2015). Differentiation of Anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate análisis. *Food Chemistry: Molecular Sciences*. 170, 243-240. <https://doi.org/10.1016/j.foodchem.2014.08.040>
- Hong, Y. (2011). NMR-based metabolomics in wine science. *Magnetic Resonance in Chemistry*, 49, S13-S21. <https://doi.org/10.1002/mrc.2832>
- Huang, F., Song, H., Guo, L., Guang, P., Yang, X., Li, L., Zhao, H. & Yang, M. (2020). Detection of adulteration in Chinese honey using NIR and ATR-FTIR spectral data fusión. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 235, 118297. <https://doi.org/10.1016/j.saa.2020.118297>
- Li, Q., Zeng, J., Lin, L., Zhang, J., Zhu, J., Yao, L., Wang, S., Yao, Z. & Wu, Z. (2020). Low risk of category misdiagnosis of rice syrup adulteration in three botanical origin honey by ATR-FTIR and general model. *Food Chemistry: Molecular Sciences*, 332, 127356. <https://doi.org/10.1016/j.foodchem.2020.127356>
- Lozano, V., Olivieri, A., Pellegrino, R & Pisano, P. (2021). Cap. 3.28 Statistics and Food Quality. En *Comprehensive Foodomics*, 03, 362-386. <https://doi.org/10.1016/B978-0-08-100596-5.22862-X>
- Maione, C., Barbosa, F. & Barbosa, R. (2019). Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. *Computers and Electronics in Agriculture*, 157, 436-446. <https://doi.org/10.1016/j.compag.2019.01.020>
- Pepi, S. & Vaccaro, C. (2017). Geochemical fingerprints of “Prosecco” wine based on major and trace elements. *Environmental Geochemistry and Health*, 40(2), pp.833-847.
- Rinnan, A°, Van den Berg, F. & Balling, S. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28(10), 1201-1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Ríos, R., Azcarate, S., Camiña, J., Callejón, R. & Amigo, J. (2019). Application of hierarchical classification models and reliability estimation by bootstrapping, for authentication and discrimination of wine vinegars by UV–vis spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 191, 42-53. <https://doi.org/10.1016/j.chemolab.2019.06.001>

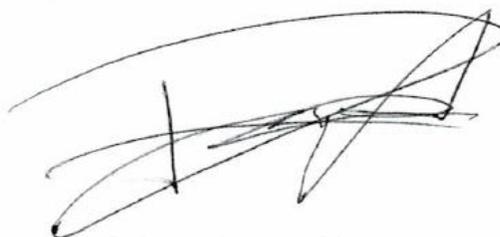
- Riswahyuli, Y., Rohman, A., Setyabudi, F., & Raharjo, S. (2020). Indonesian wild honey authenticity analysis using attenuated total reflectance-fourier transform infrared (ATR-FTIR) spectroscopy combined with multivariate statistical techniques. *Helyon. Cell Press*, 6 (4). <https://doi.org/10.1016/j.heliyon.2020.e03662>
- Rolin, D., Deborde, C., Maucourt, M., Cabasson, C., Fauvelle, F., Jacob, D., Canlet, C. & Moing, A. (2013). Cap. 1 High-Resolution 1H-NMR Spectroscopy and Beyond to Explore Plant Metabolome. En *Advances in Botanical Research* (pp. 2-36). Francia: Elsevier Ltd. <http://dx.doi.org/10.1016/B978-0-12-397922-3.00001-0>
- Sánchez, J. (2021). Perfil metabolómico basado en resonancia magnética nuclear para la clasificación de mieles mexicanas por origen floral. Tesis de Licenciatura Universidad Nacional Autónoma de México.
- Šelih, V., Šala, M. and Drgan, V. (2014). Multi-element analysis of wines by ICP-MS and ICP-OES and their classification according to geographical origin in Slovenia. *Food Chemistry*, [online] 153, pp.414-423. Available at: <https://www.sciencedirect.com/science/article/pii/S0308814613019614?via%3Dihub>.
- Tan, J., Li, R. & Jiang, Z. (2015). Chemometric classification of Chinese lager beers according to manufacturer based on data fusion of fluorescence, UV and visible spectroscopies. *Food Chemistry: Molecular Sciences*, 185, 30-36. <https://doi.org/10.1016/j.foodchem.2015.03.085>
- Ulloa, P., Guerra, R., Cavaco, A., Da Costa, A., Figueira, A. & Brigas, A. (2013). Determination of the botanical origin of honey by sensor fusion of impedance e-tongue and optical spectroscopy. *Computers and Electronics in Agriculture*, 94, 1-11. <https://doi.org/10.1016/j.compag.2013.03.001>

Vo.Bo.



Dr. Eduardo Rodríguez de San Miguel Guerrero

VoBo



Q. Pedro Anzurez Sánchez