



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS TABASCO

POSTGRADO EN PRODUCCIÓN AGROALIMENTARIA EN EL TRÓPICO

**TIIFICACIÓN DE MIEL DE DIFERENTE ORIGEN BOTÁNICO O
GEOGRÁFICO MEDIANTE LA ESPECTROSCOPIA DEL
INFRARROJO CERCANO**

CARLOS AUGUSTO REYES SÁNCHEZ

TESIS

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS

H. Cárdenas, Tabasco

2009

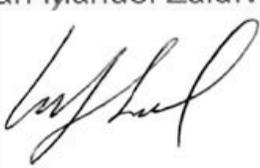
La presente tesis titulada: **Tipificación de miel de diferente origen botánico o geográfico mediante la espectroscopia de infrarrojo cercano**, realizada por el alumno Carlos Augusto Reyes Sánchez, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS
PRODUCCION AGROALIMENTARIA EN EL TROPICO

CONSEJO PARTICULAR

CONSEJERO: 
Dr. Armando Guerrero Peña

ASESOR: 
Dr. Juan Manuel Zaldívar Cruz

ASESOR: 
Dr. Joel Lara Reyna

Cárdenas, Tabasco, México, 10 de Julio de 2009

RESUMEN

Para tipificar miel existen diferentes alternativas metodológicas, en donde los procedimientos utilizados para resolver la mayoría de los problemas implican obtener información sobre los compuestos individuales de la mezcla. Cuando el problema no es qué contiene la muestra sino determinar su identidad la solución puede ser mediante señales de perfil (huellas espectrales) de un grupo de compuestos. La información contenida en las huellas espectrales se extrae mediante técnicas quimiométricas. Los métodos instrumentales modernos (espectroscopia de infrarrojo cercano, cromatografía de gases, espectrometría de masas acoplada a un generador de espacio de cabeza, espectroscopia UV-Vis, etc.), presentan gran interés principalmente por su elevada velocidad de análisis, por la reducción de los errores analíticos y en el consumo de reactivos, entre otros.

El objetivo del trabajo fue evaluar si la espectroscopia del infrarrojo cercano (NIRS) genera información suficiente para la tipificación de miel de diferente procedencia botánica o geográfica con resultados similares a los obtenidos con métodos físico-químicos. Para ello se utilizaron cuatro tipos miel (120 muestras). Los resultados obtenidos en ambos casos se trataron quimiométricamente, realizando primero el preprocesado mediante *Standard Normal Variate* (SNV) y una primera derivada, posteriormente se aplicaron técnicas para visualizar la distribución de los datos como HCA, PCA y caras de Chernoff. Posteriormente se generó el modelo de clasificación mediante LDA y el modelado de clases con SIMCA. Los modelos se validaron interna y externamente con valores de poder de clasificación del 100% y predicción mayores al 97%. Esto es, los modelos generados son altamente efectivos al clasificar y predecir una muestra problema. Por lo tanto, la tecnología de infrarrojo cercano (NIR) genera información suficiente para tipificar miel de diferente origen botánico o geográfico con resultados similares a los obtenidos en los análisis físico-químicos. El empleo de las técnicas quimiométricas es fundamental para interpretación de las señales espectrales.

ABSTRACT

In order to classify honey there are different methodological alternatives. The procedures used to solve most problems must to obtain information about the individual compounds in a mixture. When the problem is not what the sample contains but rather to determine its identity, the solution can be a signal profile (spectral fingerprint) of a group of compounds. The information contained in the spectral fingerprint is extracted by chemometric techniques. Modern instrumental methods (near infrared spectroscopy, gas chromatography-mass spectrometry coupled to headspace, UV-Vis spectroscopy, etc.), are of great interest mainly because of their high speed analysis, the reduction of analytical errors and the reduction in the consumption of reagents, among others.

The objective of this work was to evaluate whether NIRS technology generates sufficient information for the characterization of honey from different geographical and botanical origin with similar results to those obtained by physical and chemical methods. This work used four types of honey (120 samples). The results in both cases were tested with chemometric techniques, performing first the preprocessing by Standard Normal Variate (SNV) and a first derivative, then were applied to visualize the distribution of data such as HCA, PCA and Chernoff faces. Later the model of classification was generated by LDA and modeling classes by SIMCA. The two models were validated internally and externally with values of power of classification 100% and prediction higher than 97%. The models generated are highly effective to classify and predict a test sample. Therefore, NIR technology generates enough information to classify honey from different botanical or geographical origin with similar results to those obtained with the physical and chemical analysis. The use of chemometric techniques is essential for interpretation of spectral signals.

Palabras clave: Miel mexicana, Tipificación, Quimiometría, HCA, PCA.

Keywords: Mexican honey, Typifying, chemometric treatment.

Dedico esta tesis a:

A Dios porque en su infinita misericordia y paciencia, me ha prestado la vida todo este tiempo, y gracias a ello puedo lograr mis metas.

A mi esposa quien a pesar de los momentos difíciles durante este trayecto me ha dado su amor y apoyo.

A mis hijos, Haziél e Iván, quienes no han tenido un padre de tiempo completo porque les he robado del tiempo que les corresponde, pero que son la razón más importante y por la cual hago esto y tantas otras cosas.

A mis padres y hermanos quienes están conmigo siempre, sobre todo cuando más lo necesito.

A todos y cada uno de ustedes los tengo siempre presente, pues son quienes me dan la fuerza, y las ganas de seguir adelante sin importar cuán difícil sea el camino.

Agradecimientos:

Al Dr. Armando Guerrero Peña, por su amistad, apoyo y el tiempo dedicado a este proyecto.

A los Doctores Juan Manuel Zaldívar Cruz y Joel Lara Reyna, asesores e integrantes de mi Consejo Particular, por el tiempo y el apoyo brindado.

Al Dr. Wilder Camacho Chiu, porque siempre me da ánimos para seguir adelante y al cual considero un gran ejemplo y amigo.

A todos y cada uno de mis amigos, por las palabras de aliento y ánimo, y sencillamente por estar ahí cuando los he necesitado.

Al personal del Laboratorio de Suelos (LASPA) del Colegio de Postgraduados Campus Tabasco, porque durante mi estancia me brindaron su amistad y apoyo.

Un agradecimiento especial al Dr. Michele Forina, Dra. Carla Armanino, Dra. Silvia Lanteri, Dra. Mónica Casale y Dr. Paolo Oliveri, quienes me brindaron su ayuda, conocimientos y amistad durante mi estancia en el Departamento de Química y Tecnología Farmacéutica y Alimentos de la Universidad de Génova, Italia.

A los millones de mexicanos (as) que pagan impuestos, quienes, a través del Consejo Nacional de Ciencia y Tecnología (CONACYT) y el Colegio de Postgraduados, han financiado parte de mi formación.

A todos y cada uno de ustedes **MIL GRACIAS.**

CONTENIDO

INDICE DE ABREVIATURAS	ix
INDICE DE FIGURAS	x
INDICE DE CUADROS	xii
1. INTRODUCCIÓN	2
2. REVISIÓN DE LITERATURA	5
2.1. Miel.....	5
2.1.1. Composición	6
2.1.2. Tipos de miel.....	7
2.1.2.1. Miel de flores o néctar.....	7
2.1.2.1.1. Miel monoflora	7
2.1.2.1.2. Miel multiflora, mixta o poliflora.....	7
2.1.2.2. Miel de mielada	7
2.1.2.3. Miel en panal.....	8
2.1.2.4. Miel líquida.....	8
2.1.2.5. Miel cristalizada	8
2.2. Métodos de tipificación.....	8
2.2.1. Análisis palinológico.....	9
2.2.2. Análisis sensorial	11
2.2.3. Análisis físico-químicos	13
2.2.4. Métodos instrumentales modernos	13
2.3. Quimiometría.....	16
2.3.1. Reconocimiento de pautas no supervisadas	18
2.3.1.1. Análisis de agrupamientos jerárquicos	18
2.3.1.2. Análisis de componentes principales.....	22
2.3.2. Reconocimiento de pautas supervisadas	25
2.3.2.1. Análisis discriminante lineal	27
2.3.2.2. Modelado independiente de clases: "SIMCA"	30
3. OBJETIVOS E HIPÓTESIS	34
3.1. Objetivo General	34
3.1.1. Objetivos específicos	34
3.1.1.1. Realizar mediante análisis físico-químicos la tipificación de miel, con objeto de comparar los resultados obtenidos, con la tecnología NIRS.	34
3.1.1.2. Desarrollar una propuesta metodológica para la tipificación de miel mediante la tecnología NIRS.	34
3.2. Hipótesis.....	34
4. MATERIALES Y MÉTODOS	36
4.1. Descripción de espectrómetro de infrarrojo cercano (NIRS).....	36
4.2. Origen de las muestras de miel.....	37
4.3. Análisis físico-químicos	37
4.4. Análisis Quimiométrico.....	38
5. RESULTADOS Y DISCUSIÓN	41
5.1. Análisis físico-químicos	41

5.1.1.	pH.....	42
5.1.2.	Conductividad eléctrica (CE).....	44
5.1.3.	Sólidos totales disueltos (STD).....	45
5.1.4.	Grados Brix (°Brix).....	47
5.1.5.	Relación existente en los análisis físico-químicos.....	48
5.1.6.	Análisis de agrupamiento jerárquico (HCA).....	52
5.1.7.	Análisis de componentes principales (PCA).....	55
5.1.8.	Caras de Chernoff.....	56
5.2.	Análisis NIRS.....	60
5.2.1.	Análisis de agrupamiento jerárquico (HCA).....	64
5.2.2.	Análisis de componentes principales (PCA).....	67
5.2.3.	Modelos de Clasificación y Modelado de Clases.....	69
5.2.3.1.	Análisis discriminante lineal (LDA).....	70
5.2.3.2.	Modelado independiente de clases (SIMCA).....	71
6.	CONCLUSIONES	77
7.	LITERATURA CITADA	81

INDICE DE ABREVIATURAS

°Brix	Grados Brix
CE	Conductividad Eléctrica
CV	Cross Validation
HCA	Hierarchical Clustering Analysis
HMF	Hidroximetilfurfural
KNN	K - Nearest Neighbors
LDA	Lineal Discriminant Analysis
NIRS	Near Infrared Spectroscopy
PCA	Principal Components Analysis
rpm	Revoluciones por minuto
SIMCA	Soft Independent Models Class Analogy
SNV	Standard Normal Variate
STD	Sólidos Totales Disueltos
SVM	Support vector machines

INDICE DE FIGURAS

FIGURA		PAGINA
1	Distancia Euclídea y de Manhattan para las muestras k y l.	19
2	Unión de un grupo con una muestra aislada de acuerdo con los métodos del centroide (a), de la unión simple (b) y de la unión completa (c).	21
3	Dendrograma.	22
4	Análisis de componentes principales (PCA).	23
5	Representación de las puntuaciones de dos clases de objetos en los dos primeros componentes principales.	25
6	(a) Dos clases de objetos definidos por las variables x y x_2 . (b) Línea discriminante que separa las clases de objetos. (c) Proyecciones de las muestras en la recta perpendicular a la línea discriminante y cálculo de centroides.	28
7	Distancia de Mahalanobis.	29
8	Modelos de dos clases de objetos A y B en el espacio de sus componentes principales, regiones de confianza y predicción de dos muestras.	31
9	Diagrama de Coomans.	32
10	Espectrometro NIRS Versatile Agri Analyzer™ 6500.	36
11	Diagrama de caja aplicado a pH en muestras de miel.	43
12	Diagrama de caja aplicado a CE en muestras de miel.	45
13	Diagrama de caja aplicado a STD en muestras de miel.	46
14	Diagrama de caja aplicado a °BRIX en muestras de miel.	48
15	Dendrograma generado con datos de análisis físico-químicos a muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).	54

FIGURA		PAGINA
16	Puntuaciones del conjunto de muestras analizadas en los dos primeros componentes principales generados con datos de análisis físico-químicos a muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).	55
17	Representación en las caras de Chernoff de los análisis físico-químicos de las muestras de miel.	58
18	Espectros NIRS de las muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).	61
19	Regiones en espectros generados por cada una de las muestras de miel.	63
20	Dendrograma generado con las huellas espectrales NIR de las muestras de miel: multiflora Campeche (azul), Cítrico (verde), coco (café), multiflora Tabasco (rojo).	66
21	Puntuaciones del conjunto de muestras analizadas en dos componentes principales a partir de las señales espectrales NIRS de las muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).	68
22	Modelo de clasificación, LDA, obtenido con los espectros NIRS de las muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).	71
23	Gráfico del modelado independiente de clases para NIRS a muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).	74

INDICE DE CUADROS

CUADRO		PAGINA
1	Parámetros químicos de la miel de acuerdo con la NMX-F-036-1997.	6
2	Clasificación de miel según su obtención (Muselli, 2005).	9
3	Tipo y procedencia de muestras de miel.	37
4	Análisis físicos y químicos de mieles (Codex Alimentarius).	38
5	Datos promedio de análisis físicos y químicos de mieles.	49
6	Resultados LDA. Porcentaje de correcta clasificación y predicción para los cuatro tipos de miel.	70
7	Resultados SIMCA. Porcentaje de correcta clasificación y predicción para los cuatro tipos de miel.	75



1

INTRODUCCIÓN



1. INTRODUCCIÓN

Los procedimientos utilizados para resolver la mayoría de los problemas analíticos implican obtener información sobre los compuestos individuales de una mezcla. El tratamiento de la muestra incluye, una etapa de separación siendo esta la más costosa de todo el proceso. Además suele ser, también, la más larga y con mayores probabilidades de error. El procesado de los datos suele ser sencillo y la información obtenida se utiliza para detectar, identificar y cuantificar los analitos individuales.

Sin embargo, en algunas ocasiones no es necesario disponer de información individual sobre los compuestos de una muestra. El problema planteado puede resolverse satisfactoriamente con ciertas señales de perfil de un grupo de compuestos cuando el problema no es qué contiene la muestra sino determinar su identidad. La información contenida en los datos se extrae mediante la aplicación de técnicas quimiométricas. En este caso la manipulación de la muestra se reduce considerablemente, con lo que disminuyen los errores asociados a esta etapa y el tiempo total de análisis es muy corto. Ambas características ponen de manifiesto que esta estrategia de procesado matemático de la información sobre el conjunto de compuestos es muy atractiva.

El desarrollo de metodologías que permiten la reducción de la etapa de preparación de muestra es una de las tendencias actuales en Química Analítica. El objetivo es obtener la información necesaria para resolver el problema en el menor tiempo posible y reducir el costo del análisis.

Los métodos instrumentales modernos presentan, actualmente, un gran interés principalmente por su elevada velocidad de análisis. Algunas de las técnicas basadas en este tipo de generación de señales del conjunto de componentes de una muestra son: espectroscopia en el infrarrojo cercano (NIRS) (Wang, 1995;

Macho, 2002; Zhang 2002; González, 2003), e infrarrojo medio (MIR) (Carvalho, 2005), espectroscopia ultravioleta-visible (Thurston, 2004) y pirolisis acoplada a espectrometría de masas (Goodacre, 1996; Radovic, 2001; Pérès, 2002), entre otros.

Esta investigación fue realizada para evaluar si la información generada con la tecnología NIRS es suficiente para tipificar mieles de diferente origen botánico o geográfico con resultados similares a los obtenidos con los métodos físicos y químicos convencionales.

2

REVISIÓN

DE

LITERATURA

2. REVISIÓN DE LITERATURA

2.1. Miel

Miel es la sustancia dulce natural producida por las abejas a partir del néctar de las flores, de secreciones o de otras partes vivas de la planta, que las abejas recogen, transforman, combinan con sustancias específicas propias y almacenan en panales; de los cuales se extrae el producto sin la adición de ninguna sustancia (NMX-F-036-1997).

La miel, constituye el único material edulcorante que puede ser almacenado y usado tal cual es producido en la naturaleza. Para apreciar sus propiedades particulares, no requiere procesamiento o purificación alguna. Es el producto dulce viscoso elaborado por las abejas a partir del néctar de las flores. Esta definición excluye la denominada miel de mielada, que es producida por las abejas a partir de las secreciones o exudaciones de otras partes vivas de las plantas o presentes sobre ellas por la acción de ciertos insectos.

Las abejas de las especies *Apis mellifera* y *Apis dorsata* transportan, almacenan y procesan el néctar en los panales (Hooper, 1976). El procesamiento que efectúan las abejas consiste en la reducción del contenido acuoso del néctar, desde un 30-60 % a un 15-18 %; inversión de la sacarosa, por una invertasa; y la adición de glucosa-oxidasa que produce un incremento de la acidez (por la producción de ácido glucónico) y pequeñas cantidades de H₂O₂ que contribuyen junto con la baja actividad al agua a la preservación de la miel (White *et al.*, 1963).

El proceso de maduración tiene lugar dentro de las celdas abiertas del panal las cuales son selladas cuando la miel alcanza su densidad máxima. Los panales generalmente son construidos por las abejas a partir de la cera que ellas secretan. La cantidad de cera producida debe ser de 8 a 10 veces la cantidad de miel a almacenar (Hooper, 1976).

2.1.1. Composición

La miel es una solución sobresaturada de fructosa/glucosa que contiene otros carbohidratos en menor proporción como sacarosa, disacáridos reductores, oligosacáridos, dextrinas y otros componentes tales como proteínas, aminoácidos, ácidos orgánicos, sustancias minerales y varias enzimas.

En el Cuadro 1 se presentan los parámetros químicos de la miel que se deben cumplir para su comercialización con base en lo establecido en la NMX-F-036-1997.

Cuadro 1. Parámetros químicos de la miel de acuerdo con la NMX-F-036-1997.

Especificaciones	Mínimo	Máximo
Contenido aparente de azúcar reductor expresado como (g/100g) de azúcar invertido.	63.88	
Contenido de sacarosa (g/100g).	-	5
Contenido glucosa (g/100g).	-	38
Humedad (g/100g)	-	20
Sólidos insolubles en agua (g/100g)	-	0.30
Cenizas (g/100g)	-	0.60
Acidez expresada como miliequivalentes de ácido/kg.	-	40
Hidroximetilfurfural (HMF), expresado en mg/kg en miel envasada. De más de 6 meses.	-	80
Hidroximetilfurfural (HMF), expresado en mg/kg en miel envasada. De menos de 6 meses.	-	40
Índice de diastasa.*	-	8

*Para las mieles con bajo contenido enzimático, el índice mínimo de diastasa en la escala de Gothe será de 3 siempre y cuando no exceda en el contenido en HMF de 15 mg/kg.

2.1.2. Tipos de miel

La Norma Oficial Mexicana, NMX-F-036-1997, clasifica la miel en las siguientes modalidades:

2.1.2.1. Miel de flores o néctar

Es la miel que procede directamente de los néctares de las flores. Son levorrotatorias y pueden dividirse en:

2.1.2.1.1. Miel monoflora

Es aquella en cuya composición predomina el néctar de una especie vegetal determinada, entre las varias que la componen de la que toma el nombre (por ejemplo: miel de romero, azahar, lavanda, eucalipto, etc.). Debe tener las características típicas propias y además el polen de dicha especie debe ser superior al 45 % del polen total. Sin embargo, las últimas investigaciones sobre el tema, señalan que la asignación del 45 % del total de polen como criterio para definir a una miel como monoflora, debería modificarse en el caso en que las mieles provengan de plantas cuyas flores son pobres en polen (como sucede con diversas variedades de cítricos), o que poseen una biología particular, y también para aquellas plantas cuyas flores son ricas en polen como sucede con el eucalipto o el castaño (Tellería, 2001; Código Alimentario Argentino, 1999).

2.1.2.1.2. Miel multiflora, mixta o poliflora

En su composición entra el néctar de varias especies vegetales, sin que ninguna de ellas pueda considerarse predominante, es decir que ningún tipo de polen representa el 45 % total (Tellería, 2001; Código Alimentario Argentino, 1999).

2.1.2.2. Miel de mielada

Es la miel que procede principalmente de secreciones de partes vivas de las plantas.

Esta miel no procede del néctar de flores, sino de secreciones o exudaciones de otras partes vivas de las plantas o presentes sobre ella por la acción de ciertos insectos. Se denomina como miel de bosque o con el calificativo de la especie de origen (miel de encina, abeto, etc.). A diferencia de la miel de néctar de flores, es dextrorrotatoria.

2.1.2.3. Miel en panal

Es la miel que no ha sido extraída de su almacén natural de cera y puede consumirse como tal.

2.1.2.4. Miel líquida

Es la miel que ha sido extraída de los panales y que se encuentra en estado líquido, sin presentar cristales visibles.

2.1.2.5. Miel cristalizada

Es la miel que se encuentra en estado sólido o semisólido granulado como resultado del fenómeno natural de cristalización de los azúcares que la constituyen (NMX-F-036-1997).

En el Cuadro 2 se muestra la clasificación de las mieles, según su obtención propuesta por Muselli (2005).

2.2. Métodos de tipificación

El Diccionario de la Real Academia Española define tipificar como: “Ajustar varias cosas semejantes a un tipo o norma común”. Por tanto, para tipificar la miel existen diferentes alternativas metodológicas, entre las cuales podemos mencionar:

Cuadro 2. Clasificación de miel según su obtención (Muselli, 2005)

Tipo	Características
Miel escurrida	Es la miel obtenida por el escurrimiento de los panales desoperculados, sin larvas. Es de uso muy artesanal y poco frecuente.
Miel prensada	Es la obtenida prensando los panales. Es de uso poco común.
Miel centrifugada	Es el sistema mas usado actualmente. Se obtiene centrifugando los panales desoperculados.
Miel filtrada	Es la miel que se pasa por filtros de distintos tamaños para sacarle los minúsculos granos de polen y cualquier otra sustancia en suspensión y demorar de esta manera el inicio de la cristalización.
Miel pasteurizada	Es la que ha sido calentada a temperatura controlada y rápidamente enfriada para que se destruyan las levaduras, disminuyendo la formación de núcleos de cristalización, logrando que se mantenga líquida por más tiempo.
Miel orgánica	Se denomina de esta manera a la miel producida, procesada y envasada a partir de plantaciones con manejo libre de agroquímicos.

2.2.1. Análisis palinológico

Las muestras de miel son sometidas a análisis, con el siguiente procedimiento:

- Se toman 2 mL de miel, los cuales se colocan en un tubo de ensayo y se diluyen con 2 mL de agua destilada.
- Se agitan para mezclar y se centrifugan a 1500 rpm por 5 minutos.
- Con micropipeta o pipeta Pasteur graduada, se toman 0,1 mL de muestra desde el fondo del tubo, sin mover la mezcla.

-
- Esta alícuota se distribuye en cinco portaobjetos, se le agrega una gota de rojo Calberla (El rojo Calberla es un colorante especial, el cual sólo tiñe a los granos de polen dejando el resto del material (esporas de hongos, partículas de polvo, etc.) sin colorear. Y se deja secar por 5 minutos).
 - Luego se coloca encima un poco de gelatina, la cual se derrite al mechero (sin que se quemé o hierva).
 - Las preparaciones se cubren con cubreobjetos, y se dejaron secar, para luego sellarlas.
 - Estas preparaciones se analizan al microscopio óptico para identificar el polen existente, determinar a qué especie corresponden y ser contado.

Lo anterior es posible ya que el néctar recolectado por las abejas va siendo depositado en las celdillas donde completa su maduración y además, cantidades variables de granos de polen de las especies visitadas contaminan el néctar de tal manera que, un examen microscópico de él, permite identificar las especies que proveen con mayor o menor significancia néctar (Ávila *et al.*, 1993). Para esto se debe observar bajo el microscopio, absolutamente toda el área cubierta por la muestra de miel. Al contar la cantidad de granos de polen para cada especie, se calcula el porcentaje de ellas y se puede determinar a que tipo de miel corresponde la muestra (monofloral o polifloral) y con ello señalar que especies fueron elegidas por las abejas para elaborar esa miel, y junto a esto conocer sus propiedades.

Una miel que contiene 45 % o más de participación en el aporte de néctar de una especie determinada se dice que es una miel monofloral, la cual logra un mayor valor en el mercado internacional. Si no ocurre esto, por el contrario, se dice que la miel es multiflora (Ulriksen, 2003).

2.2.2. Análisis sensorial

La evaluación de los alimentos mediante análisis sensorial, es una disciplina integrada que permite establecer la calidad desde el punto de vista de los atributos del producto. Lo cual involucra la capacidad de percepción de los sentidos, y la caracterización que permita la evaluación sensorial del producto para que los catadores puedan definir fácilmente sus condiciones organolépticas. El análisis sensorial se refiere a la medición y evaluación de las propiedades de los alimentos, ingredientes o modelos evaluables por los sentidos humanos (Yáñez, 1990).

En el contexto anterior, el control de calidad de la miel se realiza considerando los atributos: 1) Apariencia (color, tamaño, forma, conformación, uniformidad) 2) Olor (los miles de compuestos aromáticos que contribuyan al aroma) 3) Sabor (dulce, amargo, salado, ácido, posiblemente astringente o metálico, entre otros) y 4) Textura (son las propiedades físicas como dureza, granulosidad, viscosidad, entre otras).

Una característica del tipo organoléptica asociada a la calidad de la miel, tiene que ver con su aspecto y la presencia de algunos componentes específicos, por esto, el color es una de las apariencias que permite juzgarla creando condiciones para la aceptación o rechazo. El parámetro por lo general se correlaciona con la madurez. La naturaleza del color, como medida discriminante del origen botánico y geográfico de la miel de *Apis mellifera*, permiten complementar otras propiedades y factores de calidad tales como el contenido de minerales, polifenoles, actividad diastásica, aminoácidos libres e hidroximetilfurfural. Este aspecto de la calidad, es el que incide de manera directa en la reacción del consumidor y es lo que se denomina calidad sensorial (Costell y Durán, 1981).

La miel multiflora, generalmente contiene dos, tres o más sabores que varían conforme al origen fitogeográfico (zona y origen floral). La interacción de estos

sabores y olores complica la evaluación sensorial del sabor en el alimento. El gusto es debido a las sensaciones que se perciben en la lengua. Este pertenece a los sentidos químicos, debido a que los receptores sensoriales, reaccionan frente a un estímulo.

Desde luego, es complejo el uso de pruebas sensoriales para establecer los atributos que contribuyen a la calidad de las mieles. Implica tiempo y trabajo, está sujeto a errores debido a la variabilidad del juicio humano y por consiguiente, es costoso. Sin embargo, no existen instrumentos que puedan duplicar o sustituir el dictamen humano.

El razonamiento para realizar un análisis sensorial se basa en los mismos principios científicos del control de proceso y variables involucradas. Por lo tanto, se hace evidente que al aplicar la metodología sea indispensable el control de las personas analistas, lugar, hora, forma y materiales adecuados para las pruebas sensoriales. Esto proporciona finalmente la confiabilidad y reproducibilidad de los atributos obtenidos, para que al final por medio de métodos estadísticos se logre un análisis confiable.

Algunas clases de compuestos orgánicos responsables del aroma en la miel, corresponden a unidades estructurales del tipo aldehídos derivados del butanal y hexanal, compuestos alcohólicos como el tetrahidrofurano, benzaldehídos y compuestos volátiles como el linalol, acetaldehídos, entre otros. Los nervios olfativos, se encargan de pasar la señal al cerebro.

Para que se perciba el olor, las sustancias deben ser volátiles, ya que ellas son transmitidas cuando el aire pasa de la nariz a la cavidad nasal, donde se encuentra el área olfatoria. Algunas clases de compuestos orgánicos responsables de olores en alimentos, son los esteroides, ácidos y aldehídos volátiles (Salamanca, 2004).

2.2.3. Análisis físico-químicos

Se entiende por análisis de miel como la actividad que conduce a la descomposición de las muestras en sus elementos físicos o químicos que las componen, con el propósito de estudiarlos o caracterizarlos en detalle.

El estudio químico de las muestra de miel puede realizarse de dos maneras: 1) para conocer “cuales” componentes se encuentran presentes, en este caso se realiza un análisis químico cualitativo, ó 2) para conocer “cuanto” de esos componentes se encuentra presente, esto es un análisis químico cuantitativo.

El análisis físico de la miel se refiere al estudio de las propiedades de los cuerpos en condiciones en que no se modifica su naturaleza íntima. Mientras que el análisis químico busca conocer la composición íntima, es decir, la clase de elementos químicos que la constituyen. La determinación de la viscosidad y el color son ejemplos de análisis físico mientras que la cuantificación del contenido de mineral y compuestos como esteres, ácidos y aldehídos son ejemplos de análisis químico (Alcantar-González *et al.*, 1992).

2.2.4. Métodos instrumentales modernos

En la industria de alimentos, dada la necesidad de conocer el origen y la calidad de los alimentos de consumo humano, es necesario realizar la determinación de ciertas propiedades que nos sirvan para caracterizarlos. Los métodos cuantitativos y cualitativos tradicionales de análisis son tediosos, demandan tiempo, mano de obra y son de un costo elevado (Murray, 1986; Osborne *et al.*, 1993; Cozzolino, 1998). Las propiedades ópticas de los alimentos y en particular la aplicación de la espectroscopia de reflectancia en el infrarrojo cercano (NIRS) se emplea desde la década del 70 en la industria alimenticia, farmacéutica, petroquímica, como alternativa a los métodos químicos y químico-biológicos convencionales. La tecnología NIRS es una técnica rápida, no destructiva, ni contaminante, de gran exactitud. En los últimos años se han desarrollado numerosas aplicaciones de la

tecnología NIRS para evaluar la composición, el procesamiento y certificación de la calidad de los alimentos. El método utiliza la región del infrarrojo cercano que comprende longitudes de onda entre los 700 y 2500 nanómetros (nm) del espectro electromagnético. En este método, cuando la luz (energía) incide sobre una muestra, una parte de los fotones se transmite a través de la misma, siendo el resto absorbida. La absorción de energía por la muestra produce que los enlaces entre C-H, N-H, O-H, S-H, C=O y C=C principales constituyentes de la estructura básica de las sustancias orgánicas; vibren en distintas formas. La interacción de la energía con la materia obedece a la ley de Lambert-Beer, que establece que la absorbancia a cualquier longitud de onda es proporcional al número o concentración de moléculas absorbentes presentes en el camino recorrido por la radiación (Cozzolino, 2002).

Esto determina que para un material de naturaleza química heterogénea el espectro obtenido en la región del infrarrojo cercano sea una combinación de bandas de absorciones parciales sobrepuestas o muy cercanas, que suelen confundirse en una línea suavizada, en que se encuentran picos, valles y curvaturas en forma de hombro (Alomar, 1998). Al desarrollar una calibración NIRS, se relaciona mediante un algoritmo la información espectral (óptica) con la información de la composición físico-química (método de referencia) a través de la aplicación de modelos estadísticos como son la regresión múltiple, los componentes principales y los mínimos cuadrados parciales.

La espectroscopia de infrarrojo cercano (NIRS) es usada en muchas áreas, los espectros NIR de una muestra son medidos típicamente por escáneres en cientos de longitudes de onda espaciados igualmente. La información en la curva es usada para predecir la composición química de la muestra por extracción de información relevante de muchos picos sobrepuestos. Aproximaciones estándar, tales como el Análisis Discriminante Lineal (LDA) han sido descritos, sin embargo, estos métodos fallan con muchas variables, así que necesitan tomarse diferentes

aproximaciones. Una solución es reducir la dimensión de la matriz de predicción y entonces aplicar LDA. Diversos métodos de clasificación se han sido comparados usando espectros de masas, incluyendo análisis discriminante lineal y cuadrático y la clasificación por el método del árbol. Y muchos autores enfatizan la necesidad de métodos para eliminar el ruido de los datos y seleccionar las características relevantes. La transformación de pequeñas ondas y el análisis discriminante estándar son aplicadas para remover el ruido.

El espectro NIR muestra sobretonos y combinación de las bandas de estos grupos. Desafortunadamente, las bandas de absorción en la región NIR son amplias y se sobreponen, lo cual significa que las técnicas convencionales de calibración univariadas, usando solo una longitud de onda por componente para evaluaciones, no pueden ser aplicadas en casos de sobreposición de bandas. El desarrollo de herramientas estadísticas más sofisticadas, como el método multivariado de regresión de mínimos cuadrados parciales (PLS), da la posibilidad para la amplia aplicación de espectroscopia NIR a muchos laboratorios (Nikolich, 2001).

La espectroscopia NIR ofrece un gran número de ventajas para análisis cualitativos y cuantitativos y aplicaciones de control de procesos. Entre las principales ventajas se tiene: no preparar la muestra para el análisis, reducción del error analítico, no hay residuos, costos reducidos, mediciones y análisis más rápidos, componentes ópticos no giroscópicos, mediciones remotas por fibra óptica, mayor exactitud en los análisis y fácil uso.

Con la técnica de reflectancia de infrarrojo cercano, la muestra ha analizar esta sujeta a radiación NIR, la superficie de la muestra absorbe parte de la energía, mientras que el resto es dispersada en todas direcciones. La luz dispersada da información sobre la composición de la muestra en la prueba y el espectro resultante es usado en el análisis cualitativo y cuantitativo (Kubelka, 1948). Los

mejores intervalos de frecuencia están en la región del espectro en el cual la intensidad de luz reflejada se correlaciona mejor con las concentraciones de las especies químicas analizadas.

2.3. Quimiometría

La Quimiometría puede definirse como una disciplina química que utiliza métodos matemáticos y estadísticos para diseñar procedimientos de medida y para obtener la mayor cantidad de información química posible a partir del análisis de datos (Massart, 1997).

Dentro de la Quimiometría pueden citarse los métodos de diseño experimental y exploración de datos, los métodos de reconocimiento de pautas, calibración y regresión simple y multivariante así como los sistemas expertos e inteligencia artificial, entre otros (Massart, 1997).

En análisis multivariante cada muestra puede considerarse como un punto en un espacio multidimensional. El objetivo es extraer la información útil de todo el conjunto de las variables medidas en una experiencia. El diseño de experimentos es un complemento de este análisis porque proporciona una estrategia para diseñar la experimentación de un proceso. Esto permite obtener la mayor información útil posible con el menor número de experimentos. Además, la calidad de los parámetros calculados dependen exclusivamente del diseño seleccionado, y son óptimos (menor varianza) cuando la matriz del diseño es ortogonal, es decir, las concentraciones de los diferentes compuestos no están correlacionadas (Massart, 1997).

Los métodos de reconocimiento de pautas tienen la finalidad de clasificar o agrupar muestras según las características de las mismas (Brereton, 1990). Pueden distinguirse dos tipos de técnicas de reconocimiento de pautas: supervisadas y no supervisadas.

Las técnicas no supervisadas tienen como objetivo determinar relaciones entre muestras sin utilizar ninguna información previa con respecto a la pertenencia a determinadas clases. Permiten conocer la existencia de estructuras en los datos y tener una idea de la utilidad de la información contenida en los mismos. Entre estos métodos se encuentran el análisis de agrupamientos, una de cuyas modalidades es el agrupamiento jerárquico (HCA) en el que las muestras se asocian en función de unos índices de similitud, y el análisis de componentes principales (PCA), en el que se reduce la dimensión de un conjunto de datos mediante el cálculo de combinaciones lineales de las variables originales.

Las técnicas de reconocimiento de pautas supervisadas tienen como fin desarrollar reglas de clasificación y/o modelado de clases para muestras (objetos) desconocidas a partir de un grupo de muestras conocidas que están caracterizadas por los valores de las variables medidas. El éxito de la predicción depende de que las reglas de clasificación y modelación así como de que los límites marcados por ellas sean adecuados. El objetivo final es predecir si una muestra desconocida pertenece a un grupo (o clase) o no. Algunas de las técnicas más importantes de clasificación y modelado de clases son el análisis discriminante lineal (LDA), modelado independiente de clases (SIMCA), “support vector machines” (SVM), el método de los k vecinos más próximos (k-NN) y las redes neuronales artificiales (ANN), entre otros.

Los métodos de calibración multivariante se basan en concentrar toda la información de las variables medidas en un número menor de ellas sin pérdida de información relevante. Dentro de estas técnicas se engloban la regresión por componentes principales (PCR) y la regresión por mínimos cuadrados parciales (PLS).

A continuación se describen brevemente los fundamentos de las diferentes técnicas quimiométricas que se utilizan en este trabajo:

- Como técnicas exploratorias HCA y PCA
- Como técnicas de clasificación y modelado de clases LDA y SIMCA

2.3.1. Reconocimiento de pautas no supervisadas

2.3.1.1. Análisis de agrupamientos jerárquicos

En esta técnica de análisis exploratorio se calcula y compara la similitud entre las muestras. Algunas de las posibilidades para determinar la similitud entre los objetos se describen a continuación (Brereton, 2003):

- Coeficientes de correlación entre las muestras.

Un coeficiente de correlación de 1 implica que las muestras tienen características idénticas. Este coeficiente se obtiene cuando una muestra se compara con ella misma. Los valores pueden oscilar entre -1 y 1 aunque es habitual trabajar con el cuadrado de este número, es decir, el cuadrado del coeficiente de correlación.

- Distancia Euclídea.

La distancia Euclídea entre dos muestras k y l está definida por:

$$d_{kl} = \sqrt{\sum_{j=1}^J (x_{kj} - x_{lj})^2} \quad (1)$$

donde J representa el número total de variables medidas y x_{kj} y x_{lj} son los valores de la variable j para las muestras k y l respectivamente. Esta medida es, en cierta forma, contraria al coeficiente de correlación y estrictamente es una medida de “no similitud”. No hay restricciones para su valor si bien es siempre positivo. La ecuación (2) es la notación matricial de la ecuación (1):

$$d_{kl} = \sqrt{(\mathbf{x}_k - \mathbf{x}_l) \cdot (\mathbf{x}_k - \mathbf{x}_l)} \quad (2)$$

- Distancia de Manhattan.

La distancia entre dos muestras k y l se determina según:

$$d_{kl} = \sum_{j=1}^J |x_{kj} - x_{lj}| \quad (3)$$

Las diferencias entre las distancias Euclídea y de Manhattan se ilustran en la Figura 1 en el caso de dos variables. Esta distancia es siempre mayor o al menos igual que la Euclídea.

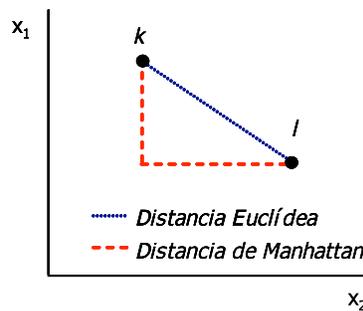


Figura 1. Distancia Euclídea y de Manhattan para las muestras k y l .

- Distancia de Mahalanobis.

Se asemeja a la distancia Euclídea pero tiene en cuenta que algunas variables pueden estar correlacionadas y por lo tanto miden en cierto grado la misma propiedad. La distancia entre dos objetos k y l se puede definir como:

$$d_{kl} = \sqrt{(\mathbf{x}_k - \bar{\mathbf{x}}_l) \mathbf{C}^{-1} (\mathbf{x}_k - \bar{\mathbf{x}}_l)} \quad (4)$$

donde \mathbf{C} es la matriz varianza-covarianza de las variables. Es una matriz simétrica cuyos elementos representan las covarianzas entre pares de variables. Los elementos de la diagonal corresponden a la varianza de cada variable. La matriz es de dimensiones $J \times J$.

Este método no es fácilmente aplicable cuando el número de variables excede al de muestras analizadas debido a que la matriz \mathbf{C} no tiene inversa. Una posible

solución requeriría seleccionar un número más pequeño de variables. Su utilización requiere trabajar con componentes principales en lugar de con las variables originales (ver sección, PCA).

El análisis de agrupamientos puede dividirse en dos tipos de técnicas: agrupamientos jerárquicos y no jerárquicos.

En los agrupamientos jerárquicos se generan dendrogramas en los que se visualizan las distancias entre muestras y en los que pueden considerarse diferentes grupos en función de determinados niveles de similitud.

En los agrupamientos no jerárquicos es necesario prefijar el número de grupos en los que se quieren dividir las muestras y éstos se generan mediante un proceso iterativo en el que la pertenencia de cada muestra a un grupo concreto se va modificando hasta llegar a una situación estable.

En el caso del análisis de agrupamientos jerarquizados (HCA), después de calcular las distancias entre todos los pares de muestras, el siguiente paso es generar los agrupamientos. Los pasos que se deben seguir son los siguientes (Brereton, 2003):

- a) Encontrar las dos muestras más similares, esto es, aquellas entre las que existe la menor distancia.
- b) Formar un grupo con estas muestras. Ahora se sustituye el grupo ya formado por un punto que lo represente. Para seleccionar al representante se suelen aplicar los métodos (Ramis, 2001) que se ilustran en la Figura 2. En el método del centroide, los puntos que ya forman grupo quedan representados por su centroide. Sin embargo en los otros métodos, el grupo queda representado por aquel punto que presente la menor distancia con respecto a la nueva muestra en el caso de la unión simple y por el de mayor distancia si se aplica el método

de la unión completa. A continuación se calculan las nuevas distancias entre el nuevo grupo formado y todas las demás muestras buscando de nuevo el valor más pequeño.



Figura 2. Unión de un grupo con una muestra aislada de acuerdo con los métodos del centroide (a), de la unión simple (b) y de la unión completa (c).

- c) Los siguientes pasos consisten en agrupar los datos de manera similar hasta que todos los objetos se unan.

El resultado es una representación gráfica llamada dendrograma. Es una estructura en forma de árbol donde la longitud de las ramas es proporcional a la distancia entre los grupos unidos. La forma del dendrograma depende del método de unión. Habitualmente, las distancias entre muestras se transforman según la siguiente escala de similitud:

$$S = 1 - \frac{d_{kl}}{d_{max}} \quad (5)$$

donde d_{max} es la mayor distancia del conjunto de datos.

En el ejemplo de la Figura 3 los objetos se dividen formando fundamentalmente 2 agrupamientos (A y B). Las muestras 1, 4, 7 y 2 presentan características similares entre sí pero distintas a las del otro grupo, constituido por las muestras 6, 8, 5 y 3.

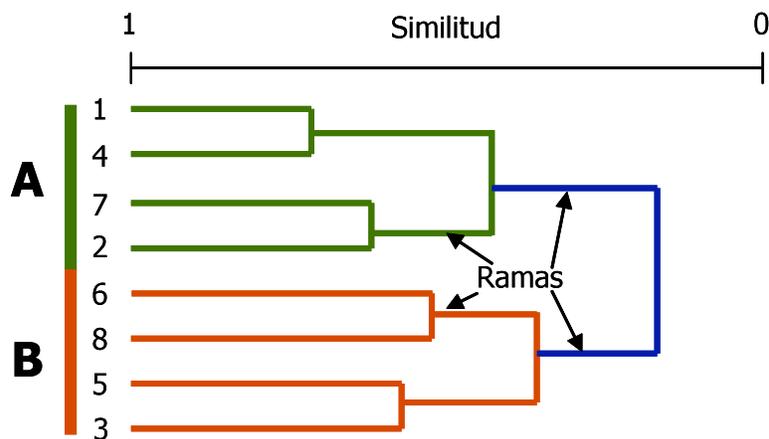


Figura 3. Dendrograma.

2.3.1.2. Análisis de componentes principales

Es una técnica de reconocimiento de pautas no supervisadas que permite encontrar la estructura “latente” en una tabla de datos cuya lectura directa lo impide. El método trata de reducir la dimensión del conjunto de datos construyendo unas nuevas variables, ortogonales entre sí, que reciben el nombre de componentes principales (PC) y que son combinaciones lineales de las variables originales.

En este análisis la matriz de datos X formada por I objetos o muestras y J variables (dimensión $I \times J$) se descompone (Massart, 1997) en el producto de las matrices T y P que contienen la información relacionada con las muestras y con las variables respectivamente.

$$X = T \cdot P + E \quad (6)$$

En la expresión (6) la matriz T contiene las puntuaciones, valores (*scores*) de cada muestra en las nuevas variables, de I objetos en A componentes principales. La matriz P contiene los pesos, coeficientes (*loadings*) en las variables originales para generar las nuevas, de J variables sobre A componentes principales. E es una matriz residual. Las matrices de puntuaciones y pesos consisten en una serie de

vectores columna y fila respectivamente según muestra la Figura 4. Ambas matrices están formadas por tantos vectores como PC puedan calcularse. Estos vectores se representan como t_a y p_a donde a es el componente considerado y reciben el nombre de vectores principales. Si la matriz de datos originales es de dimensiones $I \times J$, el número máximo de PC que pueden ser calculados es el valor de la dimensión más pequeña.

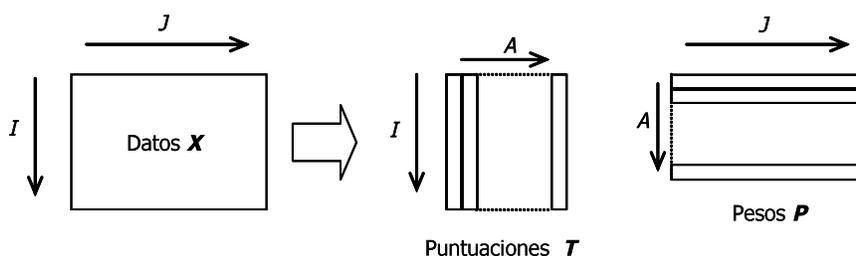


Figura 4. Análisis de componentes principales (PCA).

El objetivo de la técnica es que el número de componentes principales elegido explique la parte determinante de la matriz de datos de modo que los residuales sean lo más pequeños posibles (Blanco, 1988). En el caso de estudiar una mezcla de compuestos (Brereton, 2003), el número de PC debería ser idéntico al número de compuestos existentes. Sin embargo, esto no es generalmente tan simple y debemos tener en cuenta otros factores como por ejemplo el ruido que pueden presentar los datos. Los primeros componentes son sin duda la mayoría de las veces los más importantes. El primero explica el mayor porcentaje de varianza de los datos, el segundo el mayor de la restante y así sucesivamente. Mediante el proceso de selección del número de PC se retiene la información significativa de las matrices y se separa de aquella relacionada con el ruido.

La importancia de cada componente puede medirse mediante la suma de los cuadrados de las puntuaciones según la expresión:

$$g_a = \sum_{i=1}^I t_{ia}^2 \quad (7)$$

donde g_a es el valor propio (*eigenvalue*) del componente principal a . Además la suma de todos los valores de g_a es la suma de los cuadrados de los valores de la matriz de datos \mathbf{X} :

$$\sum_{a=1}^K g_a = \sum_{i=1}^I \sum_{j=1}^J x_{ij}^2 \quad (8)$$

donde K corresponde al valor más pequeño de I ó J . Frecuentemente estos valores se presentan como porcentajes tal y como muestra la expresión (9).

$$V_a = 100 \cdot \frac{g_a}{\sum_{i=1}^I \sum_{j=1}^J x_{ij}^2} \quad (9)$$

donde V_a es el porcentaje de información modelada por el componente a .

El porcentaje acumulado es utilizado a menudo para determinar la proporción de información que ha sido modelada.

Tras aplicar esta técnica, los datos originales quedan matemáticamente modelados por medio de unos pocos componentes principales. Usando un número A de componentes es posible establecer un modelo para cada elemento de \mathbf{X} de la forma:

$$x_{ij} = \sum t_{ia} \cdot p_{aj} + e_{ij} = \hat{x}_{ij} + e_{ij} \quad (10)$$

que es la expresión no matricial de la ecuación (6).

Una forma más clara de visualizar esta sección es buscar una interpretación gráfico-geométrica a este análisis (Blanco, 1988). Así, un objeto queda definido por un punto en el hiperespacio constituido por J variables. Esta técnica trata de reducir la dimensión construyendo unas nuevas variables de forma que los objetos

estén situados en el nuevo espacio, de forma similar a como lo estaban en el original. Esto produce un giro en el hiperespacio definido por las J variables de forma que la primera variable nueva (primer eje calculado) contenga la máxima información, la segunda (ortogonal a la primera) el máximo de la restante y así sucesivamente. La matriz de puntuaciones está constituida por las proyecciones de cada uno de los objetos del conjunto de datos en los nuevos ejes o componentes principales. La representación de las puntuaciones de los objetos sobre los componentes principales permite reconocer agrupamientos de las muestras. Esto se ilustra en la Figura 5.

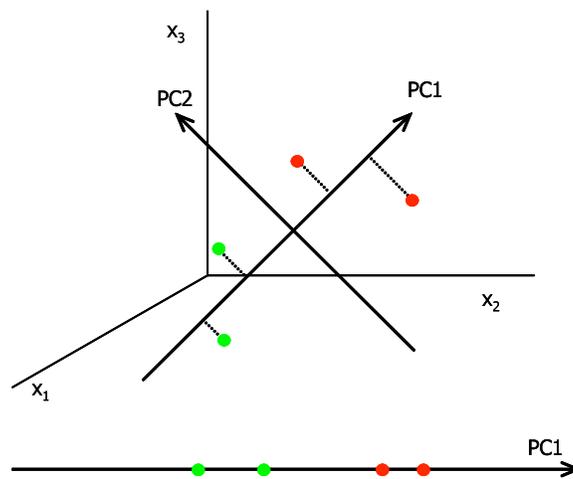


Figura 5. Representación de las puntuaciones de dos clases de objetos en los dos primeros componentes principales.

2.3.2. Reconocimiento de pautas supervisadas

Los métodos de reconocimiento de pautas supervisadas son también conocidos como métodos de clasificación. Su objetivo es clasificar muestras desconocidas en determinados grupos o clases a partir de la información suministrada por un grupo de entrenamiento. Aunque hay numerosos algoritmos es habitual utilizar una estrategia común para clasificar las muestras independientemente de la herramienta matemática que se utilice. La estrategia consta de las siguientes etapas (Brereton, 2003):

a) Crear un modelo con el grupo de entrenamiento.

El primer paso es generar un modelo matemático que relacione los valores de las variables medidas del conjunto de muestras con el grupo al que pertenecen. Estos objetos reciben el nombre de conjunto de entrenamiento.

b) Evaluación de la capacidad predictiva del modelo.

Después de generar el modelo se determina su capacidad predictiva. Un parámetro representativo puede ser el porcentaje de objetos correctamente clasificados.

Esto puede realizarse siguiendo distintos criterios. Dos de los más habituales se describen a continuación:

- Validación cruzada (*cross validation*).

Es una forma de validación interna del modelo. Una muestra o un conjunto pequeño de ellas se excluyen del conjunto de entrenamiento. Ahora se construye un nuevo modelo sin las muestras excluidas y se comprueba la capacidad de predicción del modelo sobre ellas. El procedimiento se repite hasta que todas las muestras han quedado fuera del modelo al menos una vez.

- Validación externa (*test set*).

Para comprobar la calidad del modelo se realiza la predicción de un grupo independiente de muestras que no ha sido utilizado en ninguna etapa de la construcción del modelo. Se asume que estas muestras son desconocidas y se comparan los valores obtenidos tras aplicar el modelo con los valores reales conocidos.

c) Mejoramiento de los datos.

Cuando el modelo generado no es satisfactorio existen diferentes posibilidades de mejorarlo. Por un lado se puede utilizar otro algoritmo de cálculo o realizar una

selección de variables ya que algunas pueden estar ampliamente influidas por el ruido. En lugar de utilizar todas las variables medidas se seleccionan sólo aquellas que aportan mayor información. El resto podría estar relacionada con el ruido.

d) Aplicación del modelo.

Una vez que se dispone de un modelo satisfactorio, éste podrá ser aplicado a muestras desconocidas. Un aspecto importante que se debe tener en cuenta es la estabilidad del modelo con el tiempo, es decir, cuando el conjunto desconocido de muestras se analiza semanas o incluso meses después que las correspondientes del grupo de clasificación. En algunas ocasiones es posible que existan variaciones en la respuesta instrumental registrada debido a problemas de inestabilidad de la señal con el transcurso de los días. Esto puede tener seria influencia sobre la capacidad del modelo para clasificar satisfactoriamente una muestra. En problemas de clasificación es habitual realizar algún tipo de normalización de los datos antes de su análisis.

2.3.2.1. Análisis discriminante lineal

El método más sencillo de clasificación es el univariado, en donde una variable es usada para dividir los objetos en grupos. Sin embargo, a menudo se requiere más de una medida o variable para determinar a qué grupo pertenece una muestra. En la Figura 6a los objetos representados en rojo podrían considerarse distintos de los representados en verde. Ninguna de las variables (x_1 , x_2) por separado es capaz de discriminar los objetos. Ambas son esenciales. En la Figura 6 es posible dibujar una línea (línea discriminante) entre ambos grupos. Los objetos por encima de la línea pertenecen a la clase A y los que están por debajo a la clase B. Gráficamente esto se puede representar mediante las proyecciones de los objetos sobre la línea perpendicular a la discriminante tal y como se muestra en las Figuras 6b y 6c. Ahora es fácil calcular el centro de cada clase y la distancia de una muestra desconocida a ambos centroides. La muestra desconocida se asignará al grupo más cercano.

Los resultados obtenidos con este análisis deben ser cuidadosamente estudiados ya que aquellas muestras con características de ambas clases siempre serán asignadas, exclusivamente, a una de ellas e incluso, las muestras que no pertenecen a ninguna clase siempre serán asignadas a una de ellas (Coomans *et al.*, 1979).

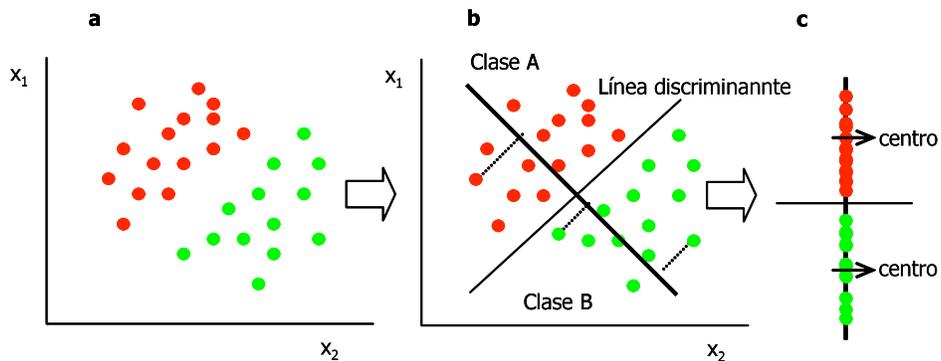


Figura 6. (a) Dos clases de objetos definidos por las variables x_1 y x_2 . (b) Línea discriminante que separa las clases de objetos. (c) Proyecciones de las muestras en la recta perpendicular a la línea discriminante y cálculo de centroides.

En muchas ocasiones se utiliza la distancia de Mahalanobis para medir distancias entre grupos de muestras. Las Figuras 7a y 7b ilustran las principales características de esta distancia (Legin *et al.*, 1997). La distancia Euclídea entre el centro del grupo (punto C) y los objetos k y h de la Figura 7a es idéntica pero k forma parte del grupo mientras que h no. Sería deseable que la distancia de C a h fuera mayor que la distancia entre C y k . La distancia de Mahalanobis indica que el punto k está más “cerca” que h de C porque está situado en la dirección del eje mayor de la elipse mientras que h no. Los objetos situados dentro de la elipse tienen valores de x_1 y x_2 que están fuertemente correlacionados. Este no es el caso de h . De manera similar en la Figura 7b, los grupos G_1 y G_2 están “más cerca” entre ellos que G_3 y G_4 según la distancia de Mahalanobis aunque presenten la misma distancia Euclídea. Todos los grupos tienen la misma forma y el mismo volumen pero G_1 y G_2 se solapan mientras que G_3 y G_4 no.

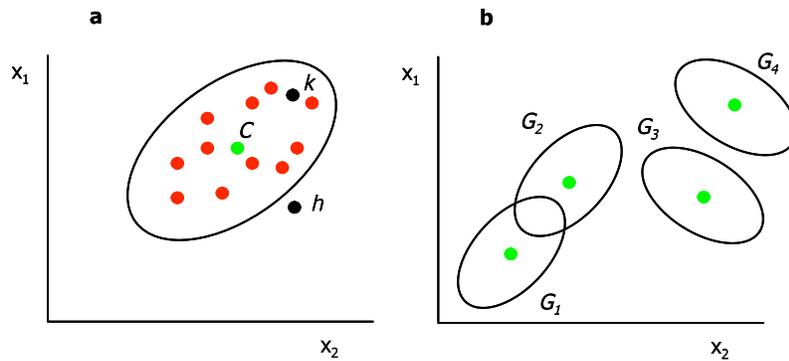


Figura 7. Distancia de Mahalanobis.

La distancia de Mahalanobis de una muestra k al centroide de una clase (A) viene definida por la expresión:

$$d_{kA} = \sqrt{(\mathbf{x}_k - \bar{\mathbf{x}}_A) \mathbf{C}_A^{-1} (\mathbf{x}_k - \bar{\mathbf{x}}_A)^T} \quad (11)$$

donde \mathbf{x}_k es el vector fila de la muestra k que recoge el valor de todas las variables y $\bar{\mathbf{x}}_A$ el que corresponde al centroide de esta clase. \mathbf{C}_A es la matriz varianza-covarianza para este grupo.

En análisis discriminante lineal pueden considerarse dos etapas: construcción de las funciones discriminantes y determinación de las reglas de clasificación. Esta técnica establece unas funciones discriminantes que forman un espacio reducido con respecto a las variables originales. La dimensión de este espacio es, generalmente, una menos que el número de clases que posea el grupo de entrenamiento. Para la construcción de las funciones se considera el caso donde existan dos clases de objetos A y B caracterizados por una serie de variables. El objetivo es ponderar y combinar linealmente las variables de manera que los conjuntos de muestras se diferencien todo lo posible reduciendo a la vez el número de dimensiones con una pérdida mínima de información (Blanco, 1988). Es posible calcular distintas funciones discriminantes lineales para separar dos clases de objetos aunque, en este caso sencillo, es suficiente una única función del tipo:

$$D = \sum_{j=1}^J w_j Y_j \quad (12)$$

donde D es la función discriminante, w_j es el coeficiente, y_j es la variable y J es el número de variables.

La obtención de los coeficientes resulta fundamental ya que no todos los ejes proporcionan una diferencia adecuada entre clases. Esto se consigue seleccionando de entre todas las posibles funciones aquellas que producen las mayores distancias de Mahalanobis entre grupos diferentes.

Si el número de clases es superior a dos se necesita un mayor número de funciones discriminantes. Éstas deben de ser ortogonales entre sí, de forma que no compartan información.

Una vez determinadas las funciones, se necesita desarrollar una regla matemática de clasificación que divida el espacio en tantas regiones como clases. Una posible regla, mencionada anteriormente, puede ser considerar la distancia de una muestra a los centros de las clases de forma que la muestra desconocida se asigna a la clase cuyo centroide esté más cercano.

2.3.2.2. Modelado independiente de clases (SIMCA)

Uno de los métodos de reconocimiento de pautas supervisadas más utilizados es el modelado de clases SIMCA (Massart, 1997; Brereton, 1990).

Mientras que en los métodos de clasificación cada muestra se asigna solamente a una clase y ninguna queda fuera de alguna de ellas, el modelado de clases permite que el resultado de la clasificación para una muestra sea que no pertenece a ninguna de las clases o que pertenece a una o más de una (con la misma o diferente probabilidad). En muchas situaciones reales no es posible

asignar una muestra concreta a alguna de las clases o existe la posibilidad de que algunas muestras pertenezcan a dos clases simultáneamente.

El método se basa en construir un modelo independiente para cada una de las clases mediante PCA. El número de componentes principales utilizados para cada clase puede fijarse si se dispone de información previa o mantener un número de ellos que explique un determinado porcentaje de la varianza de los datos. En la Figura 8 se presentan dos clases de objetos, A y B. Al aplicar PCA a la clase A (puntos verdes) se obtiene un modelo de dos componentes (un plano), mientras que la clase B (puntos rojos) se ajusta a un modelo de una única variable (una recta). Las regiones de confianza aparecen alrededor de cada clase.

La asignación de los nuevos objetos se hace comprobando si están localizados en la región de confianza de algún modelo. Así el objeto k de la Figura 8 se puede asignar a la clase B mientras que h no pertenece a ninguna clase, es un “outlier”.

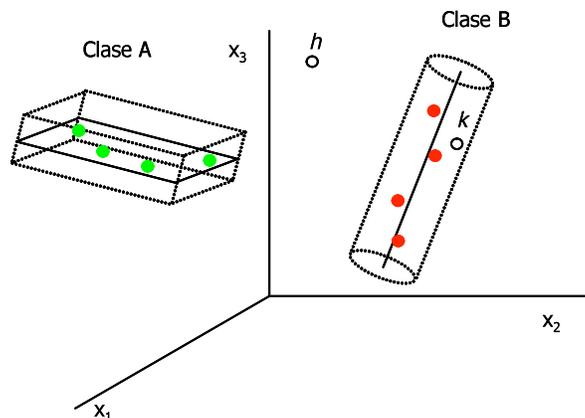


Figura 8. Modelos de dos clases de objetos A y B en el espacio de sus componentes principales, regiones de confianza y predicción de dos muestras.

Para cada clase y su modelo correspondiente se define la sensibilidad como el porcentaje de muestras que perteneciendo a esa clase son reconocidas correctamente por el modelo matemático. Análogamente se define la especificidad

como el porcentaje de muestras que perteneciendo a otra clase son reconocidas como ajenas al modelo.

Los diagramas de Coomans permiten visualizar estos índices ya que son representaciones bidimensionales de las distancias de las muestras a los modelos. La Figura 9 es un esquema de estos diagramas. En el eje de abscisas se representa la distancia a la clase A y en el de ordenadas a la clase B. La recta horizontal discontinua delimita el recinto de la clase B mientras que la línea vertical discontinua delimita el recinto de la clase A. El cuadrante inferior izquierdo (1) es el recinto común a ambas clases y cuantas menos muestras tengan sus coordenadas en él, mayor será la especificidad de los modelos. El cuadrante superior derecho (2) representa el espacio que queda fuera de ambos modelos. Las muestras con coordenadas en él son muestras extrañas a los modelos generados.

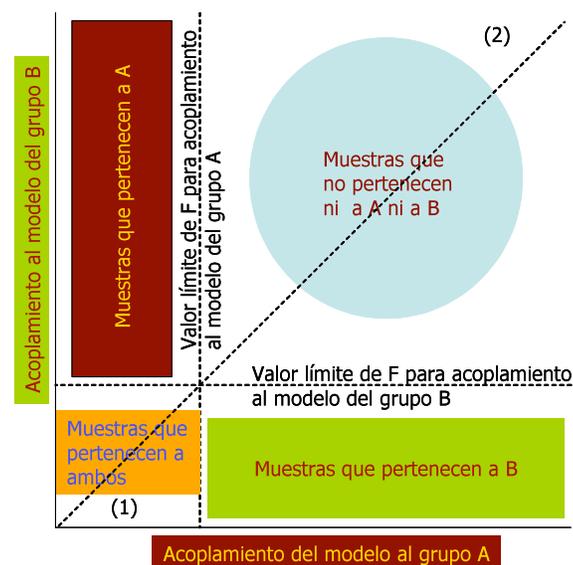


Figura 9. Diagrama de Coomans.

Cuando SIMCA se utiliza únicamente como técnica de clasificación, todas las muestras que se encuentren por encima de la diagonal serán asignadas a la clase A y las muestras por debajo a la clase B.

3

OBJETIVOS

E

HIPÓTESIS

3. OBJETIVOS E HIPÓTESIS

3.1. Objetivo General

Desarrollar una técnica rápida no destructiva basada en la tecnología NIRS para la tipificación de miel de diferente origen botánico o geográfico.

3.1.1. Objetivos específicos

- 3.1.1.1.** Realizar mediante análisis físico-químicos la tipificación de miel, con objeto de comparar los resultados obtenidos, con la tecnología NIRS.
- 3.1.1.2.** Desarrollar una propuesta metodológica para la tipificación de miel mediante la tecnología NIRS.

3.2. Hipótesis

La tecnología NIRS proporciona información útil y suficiente para la tipificación de miel de diferente origen botánico o geográfico.

4

MATERIALES

Y

MÉTODOS

4. MATERIALES Y MÉTODOS

4.1. Descripción de espectrómetro de infrarrojo cercano (NIRS)

El equipo utilizado es un espectrómetro de Reflectancia en el Infrarrojo Cercano modelo Versatile Agri Analyzer™ 6500 System II Vertical Type 461008, Serial No. 9503 100-240 Vac max 200 W, 50-60 Hz de la empresa FOSS NIRSystems (Figura 10), el intervalo que cubre el escáner es de 400 a 2500 nm en longitudes de onda pares.



Figura 10. Espectrómetro NIR Versatile Agri Analyzer™ 6500.

Para realizar la lectura el equipo cuenta con una lámpara de 4 volts, ajustada a una placa prealineada, que manda el haz de luz blanca la cual contiene todas las frecuencias al *Grating*, una superficie holográfica de 570 líneas por mm y una distancia de 1.75 μm entre líneas, el cual desvía la luz en ángulos diferentes dependiendo de la longitud de onda, ésta luz reflejada es monocromada y el ángulo de inclinación del *Grating* es controlado por la tarjeta madre, al filtro *Order sorter*, que esta sincronizado con el movimiento del *Grating* y es movido por medio de un motor de pasos, permite el paso de la luz por longitud de onda para transmitirla a la muestra, contenida en una copa o celda opaca de cuarzo, y esta a

los detectores de Silicio (400 a 1100 nm) y PbS (1100 a 2500 nm), que registran la cantidad de luz reflejada y/o transmitida, para transformarla en una señal que se envía mediante el programa ISIScan (versión 2.81 Copyright 2006) a una computadora (Acer AMD Sempron™ Processor 3200+ 1.81 GHz, 448 MB de RAM).

4.2. Origen de las muestras de miel.

Para el desarrollo de esta investigación en la que se va probar que la tecnología NIRS y la aplicación de algunas técnicas quimiométricas nos permiten lograr la diferenciación y clasificación de diferentes tipos de miel, así como, comparar los resultados obtenidos con la metodología convencional, es decir, mediante aplicación de técnicas analíticas experimentales referenciadas en el *Codex Alimentarius*, se contó con 120 muestras de miel que difieren en su origen geográfico y botánico según se muestra en el Cuadro 3.

Cuadro 3. Tipo y procedencia de muestras de miel.

Muestras	Tipo	Procedencia
30	multiflora	CP-Campus Campeche, Campeche, Mex
30	multiflora	CP-Campus Tabasco, Tabasco, Mex
30	coco	Paraíso, Tabasco, Mex
30	cítrico	Huimanguillo, Tabasco, Mex

4.3. Análisis físico-químicos

Los análisis realizados como alternativa metodológica para caracterizar miel y comprobar la capacidad de la metodología obtenida con las señales generadas con la tecnología NIRS y la quimiometría, se muestran en el Cuadro 4 en donde se indican las mediciones físico-químicas descritas en el *Codex Alimentarius*.

Cuadro 4. Análisis físicos y químicos de miel.

Análisis	Medición
pH	Potenciometría
Conductividad eléctrica (CE)	Conductimetría
Sólidos totales disueltos (STD)	Conductimetría
Brix	Refractometría

(Codex Alimentarius, 2001)

4.4. Análisis Quimiométrico

Los datos obtenidos como resultado de los análisis físico-químicos y de las lecturas en el espectrómetro se analizaron quimiométricamente, utilizando el programa Pirouette Chemometrics Software versión 3.11 para Windows (Infometrix Inc. Copyright 2007), aplicando las técnicas de reconocimiento no supervisadas para determinar relaciones entre muestras sin utilizar ninguna información previa con respecto a la pertenencia a determinadas clases. Permitiendo conocer la existencia de estructuras en los datos y tener una idea de la utilidad de la información contenida en los mismos. Se aplicó análisis de agrupamiento jerárquico (HCA) en el que las muestras se asocian en función de índices de similitud, y análisis de componentes principales (PCA) que permite encontrar la estructura “latente” en el cuadro de datos cuya lectura directa lo impide; reduciendo la dimensión del conjunto de datos construyendo unas nuevas variables, ortogonales entre sí, que son combinaciones lineales de las variables originales.

Así mismo, se aplicaron técnicas de reconocimiento de pautas supervisadas, para clasificación, LDA, que nos permite la creación de una regla de clasificación entre clases; y para modelado de clases, SIMCA, que nos permite generar un modelo independiente para cada una de las clases mediante PCA. Disponiendo de información previa como el número de componentes principales utilizados para

cada clase o manteniendo el número de ellos que contengan la máxima información discriminante.

5

RESULTADOS

Y

DISCUSIÓN

5. RESULTADOS Y DISCUSIÓN

La presentación de los resultados y su discusión se realiza con la secuencia establecida en los objetivos. Primero se exponen los resultados del análisis físico-químico con objeto de tipificar la miel, ya que debe comprobarse que son de diferente origen a fin de que puedan ser utilizadas como un referente para comprobar la hipótesis de que la tecnología NIRS es capaz de generar información útil y suficiente para tipificar la miel.

5.1. Análisis físico-químicos

Los análisis realizados fueron pH, conductividad eléctrica (CE), sólidos totales disueltos (STD) y grados Brix ($^{\circ}$ Brix).

Para la presentación y discusión de estos resultados se utilizaron los diagramas de caja o Boxplot, así como los datos promedio, su desviación estándar y su coeficiente de variación; y finalmente un análisis mediante técnicas quimiométricas no supervisadas.

Para la interpretación del Boxplot se deben tener en cuenta las siguientes consideraciones (Tukey, 1977):

- Mientras más larga la caja y los bigotes, más dispersa es la distribución de datos.
- La distancia entre las cinco medidas descritas en el Boxplot puede variar, sin embargo, la cantidad de elementos entre una y otra es aproximadamente la misma. Entre el límite inferior y el primer cuartil (Q_1) hay igual cantidad de opiniones que de (Q_1) a la mediana, de ésta al tercer cuartil (Q_3) y de (Q_3) al límite superior. Se considera aproximado porque

podiera haber valores atípicos, en cuyo caso la cantidad de elementos se ve levemente modificada.

- La línea que representa la mediana indica la simetría. Si está relativamente en el centro de la caja la distribución es simétrica. Si por el contrario se acerca al primer o tercer cuartil, la distribución pudiera ser sesgada a la derecha (asimétrica positiva) o sesgada a la izquierda (asimétrica negativa).
- La mediana puede inclusive coincidir con los cuartiles o con los límites de los bigotes. Esto sucede cuando se concentran muchos datos en un mismo punto. Pudiera ser este un caso particular de una distribución sesgada o el caso de una distribución muy homogénea.
- Cuando los bigotes de una caja presentan un traslape con los bigotes de otra caja, al realizar la proyección de los mismos, en el eje correspondiente, se dice que no existe diferencia significativa. Y cuando no existe dicho traslape entonces se puede decir que existe diferencia significativa entre clases.

5.1.1. pH

Cuando se realizó la construcción del Boxplot (Figura 11), para la determinación de pH se encontró en primer término aquellas muestras cuyos datos presentaron mayor dispersión y son los de la miel multiflora de Tabasco lo cual se refleja en la amplitud de la caja y de los extremos; en segundo término se encontró la miel de cítrico, la cual presenta gran dispersión reflejada en la amplitud de los extremos pero menor dispersión en el intervalo intercuartílico es decir, en la caja; en tercer lugar la miel procedente de Campeche, presenta una dispersión similar a la de cítrico en los extremos pero mayor en el intercuartílico y también se puede observar que es la miel en la cual se presentan los valores de pH más bajos, por tanto, este tipo de miel presenta una diferencia muy significativa con respecto a las

otras mieles analizadas; y por último, en la miel de coco se observa mayor homogeneidad en los datos ya que es la de menor amplitud en los extremos así como en la caja.

Así mismo, en la Figura 11 podemos observar, en general como las distribuciones presentan un sesgo negativo ya que la mediana en todos los casos se encuentra cerca del tercer cuartil. Lo cual nos indica que los valores de mayor frecuencia se encuentran ubicados a la derecha o hacia los valores altos de cada lote de muestras.

Cabe señalar que si nosotros quisiéramos separar entre los diferentes tipos de miel analizadas únicamente se presentaría diferencia significativa en el caso de la miel de Campeche con respecto a las otras mieles, pero en el caso de las demás sería más complicado pues estas no presentan diferencia significativa entre ellas.

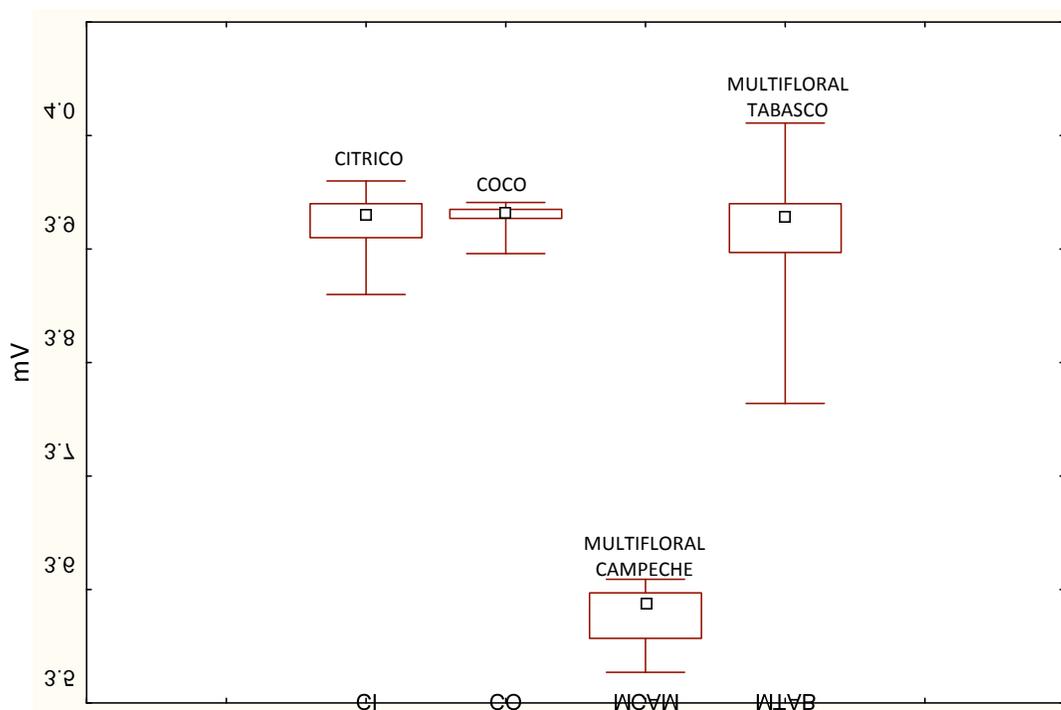


Figura 11. Diagrama de caja aplicado a pH en muestras de miel.

5.1.2. Conductividad eléctrica (CE)

En el caso de la determinación de CE al realizar la construcción del diagrama de caja se puede observar, en la Figura 12, que al igual que el resultado del análisis de pH, la miel de tipo multiflora de Tabasco muestra una gran dispersión entre muestras sobre todo después del tercer cuartil pues la línea superior se observa muy prolongada, sin embargo se puede observar que el grupo de datos presenta alguna homogeneidad pues el tamaño de la caja es muy reducido, además de presentar un sesgo positivo indicándonos que los datos de mayor frecuencia se ubican en valores bajos. En segundo término encontramos la miel de coco la cual muestra la mayor dispersión dentro del primer cuartil, y también presenta la mayor dispersión de datos pues es la que presenta el mayor tamaño de la caja, además presenta un sesgo negativo indicando que sus valores tienden a ser altos.

En el caso de la miel de cítricos en ésta se puede observar que aunque presenta cierta dispersión entre datos esta no es significativa pues su intervalo es reducido, además de presentar un sesgo ligeramente negativo indicando que los valores se encuentran no muy lejos del centro de la normal. Sin embargo, se puede observar en la Figura 12 la miel multiflora de Campeche, que presenta menor dispersión pues la caja y los bigotes están muy próximos; además de mostrar un sesgo positivo ligeramente apreciable indicando que los valores de CE son bajos, y nuevamente es la miel que presenta diferencia significativa con respecto a las otras mieles.

En el caso de querer diferenciar los diferentes tipos de miel, con base en los valores de CE podemos observar que se aprecia diferencia significativa de la miel de Campeche con respecto a las otras mieles analizadas y en el caso de la miel de cítrico esta presenta diferencia significativa únicamente con las mieles de coco y multiflora de Campeche pero no con respecto a la multiflora de Tabasco. Una diferencia significativa se observa para la miel de coco con respecto a las mieles

de cítrico y multiflora de Campeche pero no con respecto a la multiflora de Tabasco.

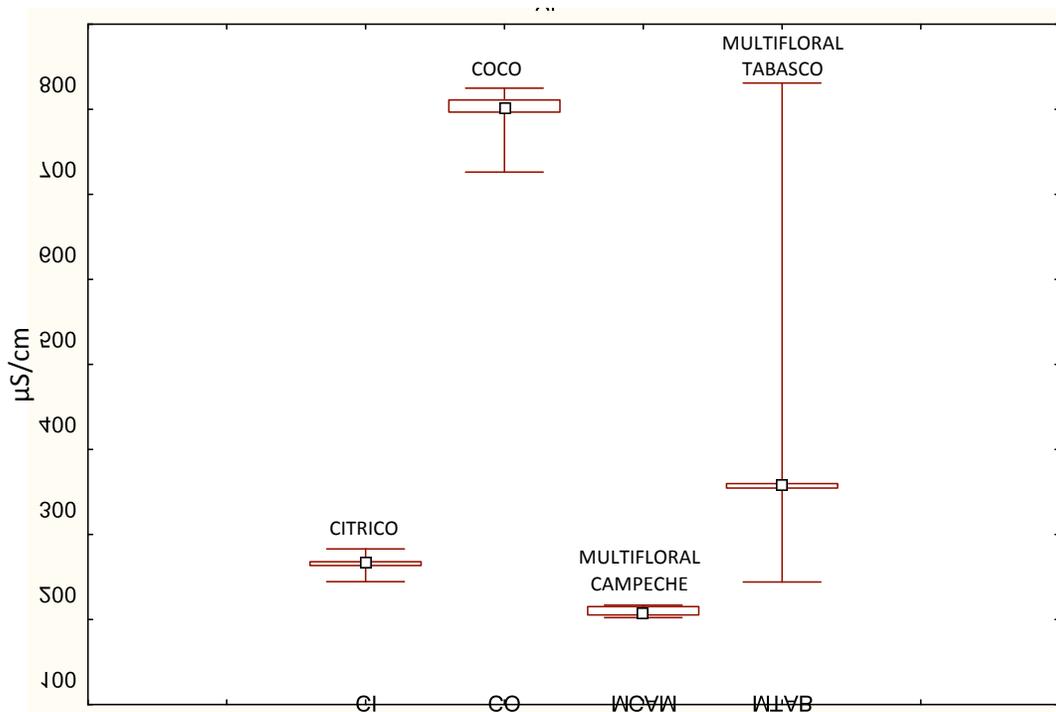


Figura 12. Diagrama de caja aplicado a CE en muestras de miel.

5.1.3. Sólidos totales disueltos (STD)

Para los resultados obtenidos del análisis de STD al realizar la construcción del diagrama de caja, como se muestra en la Figura 13, se puede observar que el comportamiento es muy similar al de la CE. La miel multiflora de Tabasco es el grupo que presenta la mayor dispersión en los datos sobre todo después del tercer cuartil lo que nos presenta un sesgo positivo indicando que los datos se ubican a la izquierda de la mediana y siendo estos de valor bajo, también se puede observar que en el caso de la miel de cítrico presenta una dispersión ligeramente mayor a la observada con respecto al análisis anterior, y con un sesgo ligeramente negativo. Pues los valores de mayor frecuencia se localizan a la derecha de la mediana.

En el caso de la miel de coco en ésta se presenta una dispersión más apreciable y con un sesgo negativo sobretodo se puede observar que la mayor dispersión se presenta en el primer cuartil. Además de ser la que presenta los valores más altos en contenido de STD.

Para el caso de la miel de Campeche se puede observar que ésta es la que presenta la menor dispersión de los datos además de que se puede observar sesgo positivo en la distribución de los mismos.

Así mismo, como se puede observar en la Figura 13, el grupo de muestras de la miel multiflora de Campeche es la única que presenta diferencia significativa con respecto a las otras mieles analizadas, en el caso de la miel de cítrico ésta presenta diferencia significativa con respecto a la miel de coco y multiflora de Campeche pero no con la multiflora de Tabasco, la miel de coco presenta diferencia significativa con respecto a la miel de cítricos y la multiflora de Campeche pero no con la multiflora de Tabasco.

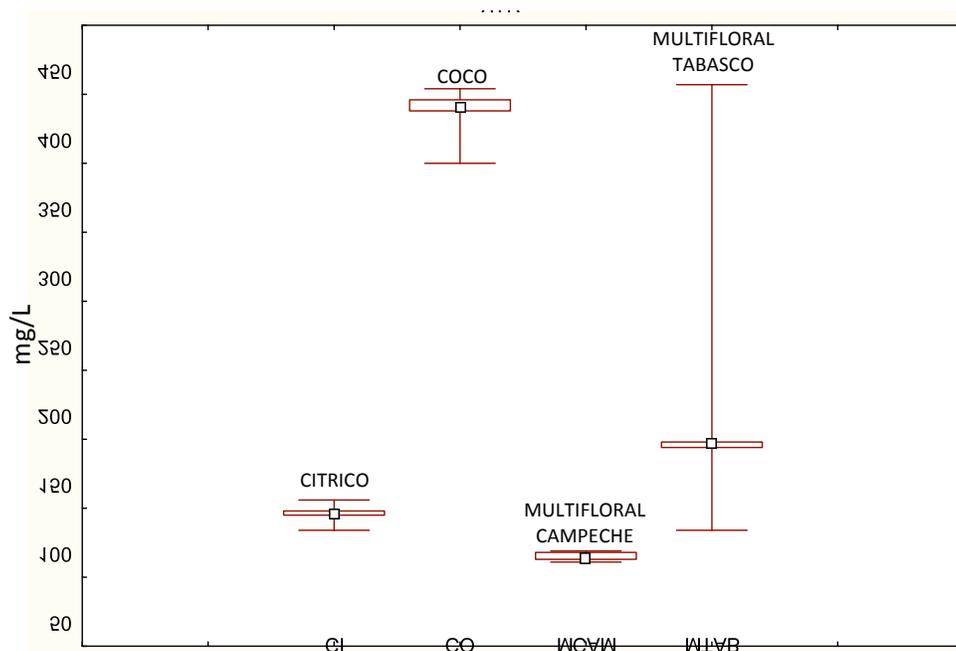


Figura 13. Diagrama de caja aplicado a STD en muestras de miel.

5.1.4. Grados Brix (°Brix)

Como se puede observar en la Figura 14, al realizar el diagrama de caja para los datos obtenidos del análisis de °Brix, se observa que en el caso de las mieles de cítricos y multiflora de Tabasco presentan igual dispersión de los datos; sin embargo en el caso de la miel multiflora de Tabasco ésta presenta una mayor dispersión intercuartílica y un sesgo positivo, encontrándose que la mayoría de los datos son bajos, la mediana se encuentra en el centro de la caja lo que nos indica una distribución simétrica de los datos. En el caso de la miel de cítricos su dispersión es menor y su sesgo es negativo con valores altos, la mediana para estas muestras coinciden con el tercer cuartil, además de ser la que presenta los valores más altos en °Brix, para la miel de coco podemos observar que ésta presenta los valores más bajos en °Brix con respecto a las otras mieles analizadas, además de presentar un sesgo positivo. La mediana para estas muestras coincide con el primer cuartil en el caso de la miel multiflora de Campeche, esta presenta una dispersión intercuartílica y un sesgo positivo, igual al observado en la miel de multiflora de Tabasco con la diferencia de que la mediana coincide con el tercer cuartil.

Cabe señalar que la única miel que presenta diferencia significativa es la miel de coco con respecto a la miel de cítrico y la multiflora de Campeche; en el caso de las otras mieles no muestran diferencias significativas entre si.

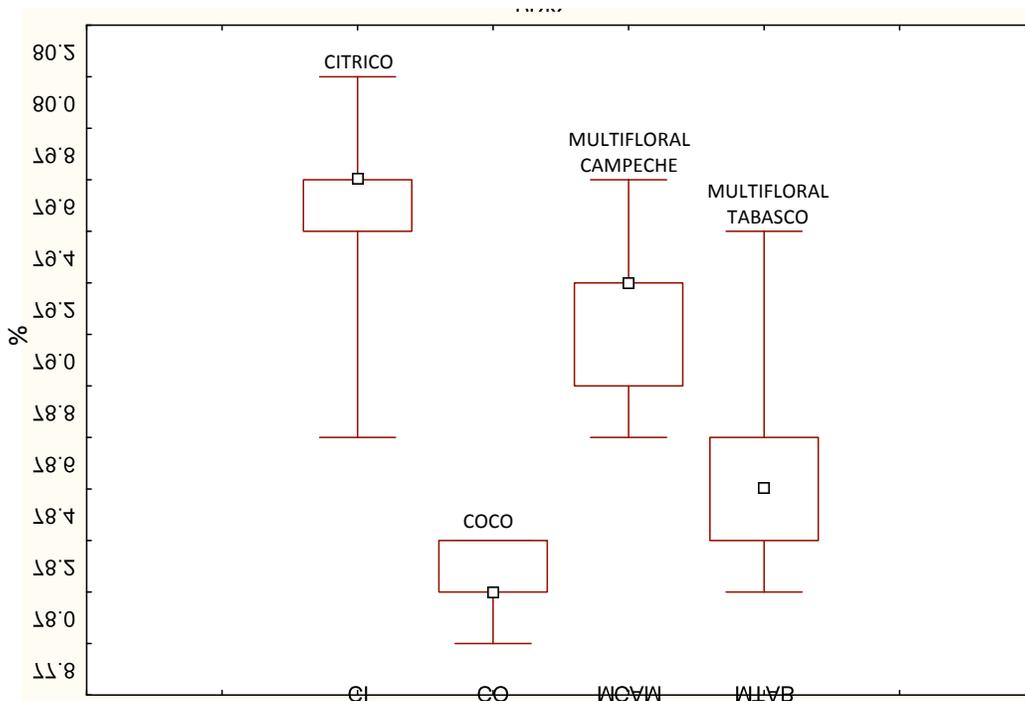


Figura 14. Diagrama de caja aplicado a °BRIX en muestras de miel.

5.1.5. Relación existente en los análisis físico-químicos.

En el Cuadro 5 se presentan los valores promedio con su desviación estándar de los datos obtenidos como resultado de los análisis físicos y químicos (pH, CE, STD, °Brix) de las muestras de miel.

Estos valores fueron utilizados en la generación de los diagramas de caja en donde se puede observar que no existe una variabilidad o diferencia significativa entre lotes de muestras por lo que no serian de utilidad si se consideran cada uno por separado para diferenciar las muestras de miel; a partir de estos datos se dedujo el coeficiente de variación el cual encontramos que oscila entre 0 y 2.8% excepto en las variables CE y STD en la miel de tipo multiflora perteneciente al estado de Tabasco ya que los datos obtenidos en estos análisis presentan un coeficiente de variación de hasta 32.4% lo que nos indica la mayor variabilidad

existente entre muestras de este lote, y lo mismo pudo ser observado en las Figuras 12 y 13, respectivamente.

Cuadro 5. Datos promedio de análisis físicos y químicos de mieles.

MIEL	pH	CE	STD	°BRIX
	$\bar{x} \pm s$	$\bar{x} \pm s$	$\bar{x} \pm s$	$\bar{x} \pm s$
CITRICO	3.92 ± 0.023	264.92 ± 7.332	145.60 ± 4.082	79.73 ± 0.231
COCO	3.93 ± 0.012	799.30 ± 19.49	439.70 ± 10.65	78.25 ± 0.101
MULTIFCAMP	3.58 ± 0.025	209.37 ± 5.117	115.03 ± 2.846	79.29 ± 0.250
MULTIFTAB	3.91 ± 0.058	377.46 ± 22.28	207.63 ± 67.28	78.71 ± 0.388

Al realizar el análisis de los datos obtenidos se encontró que existe una relación directa o inversamente proporcional entre algunos de ellos, es decir, que a partir de una de las variables podemos estimar otra como es el caso entre las variables CE y STD lo cual nos permite proponer una ecuación que nos permita estimar una variable a partir de otra, y con esto obtener un ahorro en tiempo y costo de análisis, así pues para este caso específico se encontró que si se desea conocer el valor correspondiente de STD este se puede estimar por medio de una ecuación de regresión (13) conociendo únicamente el valor correspondiente de CE:

$$STD = 0.5503 * CE - 0.1604 \quad (13)$$

Esto es, que a partir de datos conocidos de CE mediante esta ecuación obtenida se puede estimar el valor de STD para cualesquier muestras de miel, pues esta ecuación nos presenta un valor de $R^2=1$, es decir, que nos estima el valor de la variable con un 100 por ciento de confiabilidad para la variable buscada.

Por el contrario si se desea conocer el valor correspondiente a CE, este se puede estimar a partir de datos de STD presentes en las muestras de miel mediante la ecuación de regresión (14):

$$CE = 1.817 * STD + 0.294 \quad (14)$$

Esto significa que a partir de valores de contenido de STD, se puede estimar el valor de CE para cualesquier muestras de miel a partir de los valores de STD pues esta ecuación presenta un valor de $R^2=1$, es decir, que la correlación entre las variables es del 100%.

También se encontró que a partir de los datos obtenidos de °Brix y teniendo como referencia que el refractómetro nos proporciona la lectura en índice de refracción o en porcentaje de sólidos o en este caso de azúcar o presente en una solución azucarada, y que por lo tanto, el resto es agua expresada en porcentaje de humedad tenemos que para conocer el valor de la humedad en miel se establece la ecuación (15):

$$\% Humedad = -1 * Brix + 100 \quad (15)$$

Cabe señalar que la mayoría de los valores obtenidos en las determinaciones practicadas a las muestras de miel se encuentran dentro de los intervalos establecidos en las referencias bibliográficas por ejemplo en el caso del pH debe encontrarse entre valores 3.4 y 6.1 (*National Honey Board*) y en el caso de las muestras analizadas tenemos que el pH obtenido se encuentra entre los valores de 3.58 y 3.93 mV, en el caso de CE de acuerdo con (Bogdanov, 1999) esta debe encontrarse en valores menores o iguales a 0.800 mS/cm y en el caso de las mieles analizadas tenemos valores que van desde los 0.207 hasta los 0.799 mS/cm; así mismo, en el caso de la cantidad de STD tenemos que de acuerdo con la norma NMX-F-036-1997, la cantidad de STD debe encontrarse en valores no mayores a 0.3 %, sin embargo tenemos que aunque la miel de cítricos, multiflora de Tabasco y multiflora de Campeche se ajustan, en el caso de la miel de coco esta se encuentra por encima de este valor establecido pues presenta valores de 0.42 a 0.45 %.

Por ultimo en el caso de los azúcares tenemos que de acuerdo con la norma NMX-F-036-1997 la cantidad de azúcares expresados como % de azúcar invertido debe tener un mínimo de 63.88% y en el caso las mieles analizadas al realizarle la medición de la cantidad de azucars en % de azúcar (°Brix) ésta presenta valores desde 78.0 a 80.2 °Brix, y dado que la mayoría de los azucars presentes en la miel son fructosa y glucosa (Bogdanov, 1999), y estos representan el 90% del contenido de azucars reductores, entonces tenemos que la cantidad mínima establecida es similar a la presente en el caso de las mieles analizadas.

En el caso de la cantidad de cenizas Piazza, (1991) encontró que existe una relación lineal entre la conductividad eléctrica (CE) y la cantidad de cenizas en muestras de miel por lo tanto la cantidad de cenizas puede ser estimada a través de la ecuación:

$$\% \text{ Cenizas} = (CE * 0.14) / 1.74 \quad (16)$$

Y al aplicar la ecuación a los datos obtenidos de CE en la muestras de miel tenemos que los valores en % de cenizas se encuentra entre valores de 0.0806 y 0.0809% valores muy por debajo del establecido por la norma NMX-F-036-1997 de 0.6%.

Así pues una vez analizados los datos obtenidos, con métodos estadísticos descriptivos, lo cual nos da un panorama más amplio y nos permite conocer mejor las muestras que se incluyeron en la investigación; se procedió a realizar el análisis quimiométrico a dichas muestras, con objeto de poner de manifiesto las estructuras o pautas entre las muestras, y cuyos resultados se presenta en las siguientes secciones de este capítulo iniciando con el análisis de agrupamiento jerárquico.

5.1.6. Análisis de agrupamiento jerárquico (HCA)

Antes de realizar el tratamiento quimiométrico para la creación de las reglas de clasificación y el modelado de clases, se aplicaron técnicas no supervisadas para visualizar el comportamiento de los datos, primero se llevó a cabo el análisis de agrupamientos mediante el programa Pirouette, generando un diagrama de árbol o dendrograma correspondiente, que nos permite visualizar mejor las diferencias o similitudes en nuestra población de datos, como se puede observar en la Figura 15.

Como se puede observar, en la Figura 15, en el dendrograma se generan cuatro grupos, a los cuales para efectos de visualización de cada tipo de miel se les asignó un color, multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo). Puede observarse que en el caso de la miel de tipo multiflora, procedente del estado de Campeche, presenta cierta variabilidad interna aunque no significativa, y esta se ve reflejada dentro del grupo formado, como dos agrupamientos internos, esto es, a pesar de existir diferencias entre las muestras de este tipo de miel, dicha diferencia no influye de manera considerable por lo que dichas muestras forman un grupo separado del resto de las muestras analizadas, así mismo, es importante mencionar que este grupo presenta un índice de similitud igual a 0.811, un dato que nos indica la similitud de las muestras.

En el caso de la miel de cítrico procedente del estado de Tabasco, se observa que el agrupamiento es notoriamente diferente al anterior ya que en este caso aunque se obtiene el grupo esperado entre muestras representativas de este tipo, este grupo presenta internamente cuatro subgrupos, mostrando la variación presente en este grupo de muestras. Sin embargo, a pesar de presentar esta variabilidad, se observa un grupo bien formado el cual tiene un índice de similitud igual a 0.802.

La miel de coco procedente del estado de Tabasco como puede verse es el grupo mejor formado ya que aunque al igual que los otros grupos muestra también variabilidad entre muestras, esta variabilidad es menor por lo que como resultado se forma el grupo correspondiente a esta clase, cuyo índice de similaridad es de 0.860, que es el mayor valor presentando por los diferentes grupos de muestras.

Sin embargo, en el caso de la miel de tipo multiflora procedente del estado de Tabasco, por presentar menor homogeneidad afecto la formación del grupo, por lo que se forman dos agrupamientos, uno de los cuales se acopla con el grupo muestras de cítrico, y otro pequeño grupo con las muestras de coco.

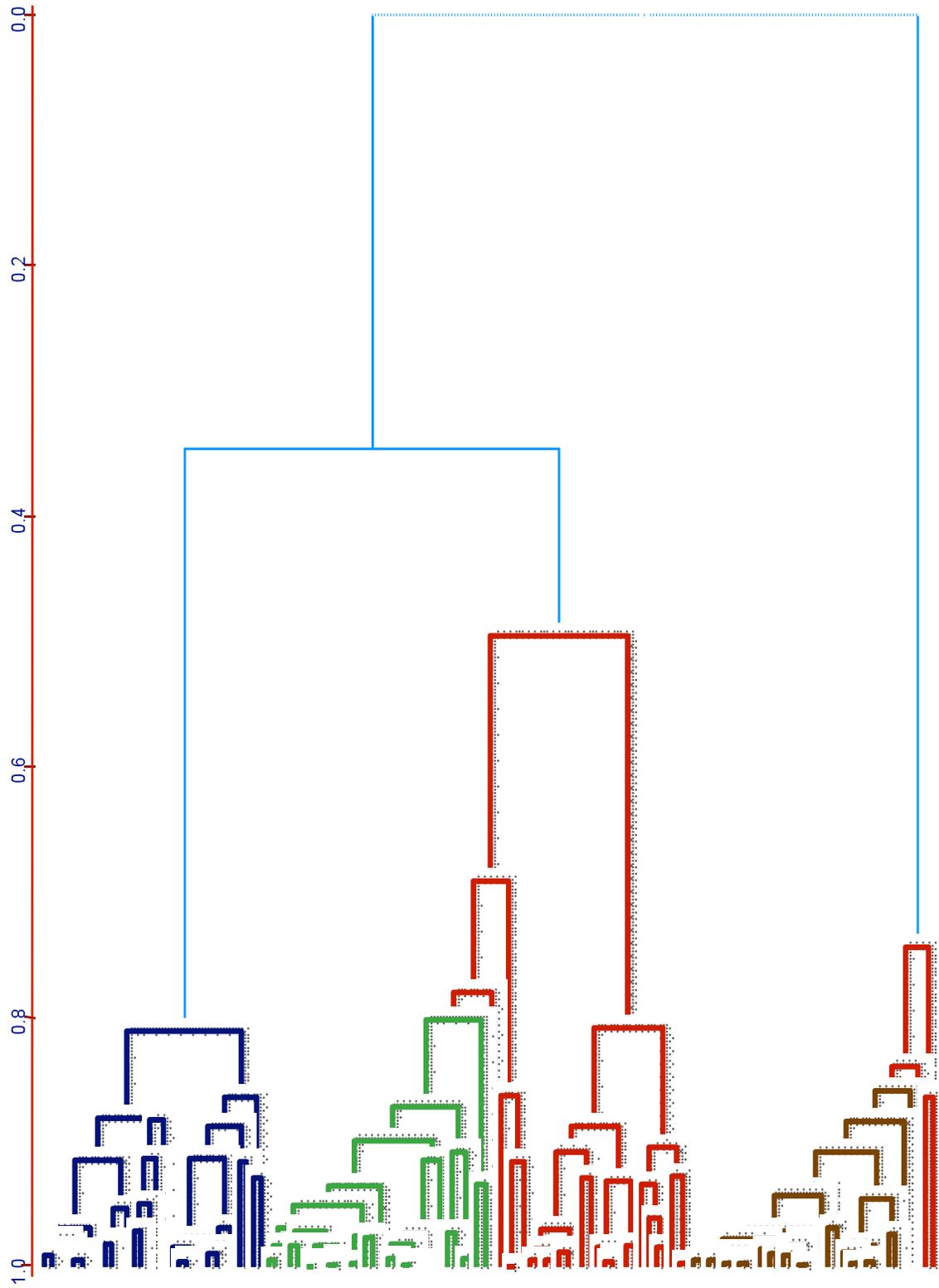


Figura 15. Dendrograma generado con datos de análisis fisico-químicos a muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).

5.1.7. Análisis de componentes principales (PCA)

Otra técnica que se aplicó a los datos es el análisis de componentes principales la cual a diferencia de la anterior que se basa en la medición de la distancia, en este caso la distancia Euclídea, entre muestras. PCA realiza una rotación ortogonal de los ejes de manera que las muestras se presenten en un ángulo diferente para lo cual se crean nuevas variables que son una combinación lineal a partir de las variables originales, dicho análisis se realizó mediante el programa Pirouette, para posteriormente mediante la selección de los componentes significativos generar un gráfico que nos permite apreciar la distribución de los datos en el espacio de dichos componentes, como se muestra en la Figura 16.

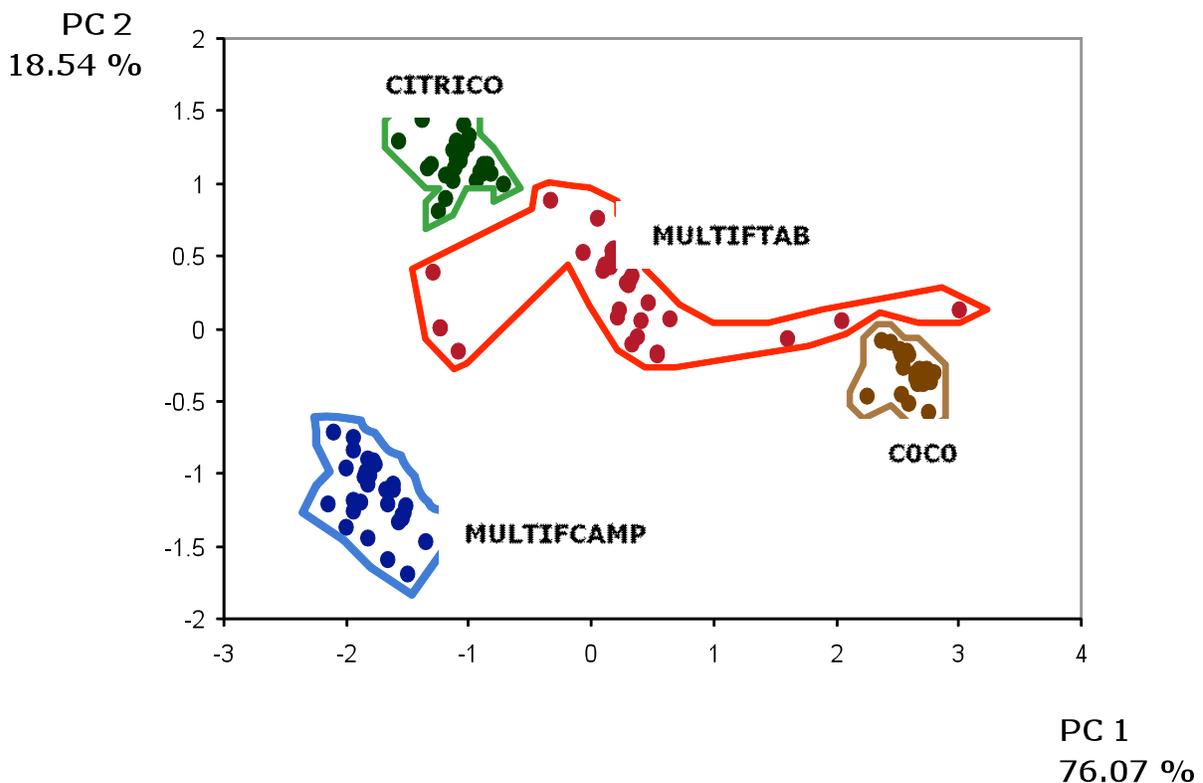


Figura 16. Puntuaciones del conjunto de muestras analizadas en los dos primeros componentes principales. Generados con datos de análisis físico-químicos a muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).

En la Figura 16, se graficaron las puntuaciones de los primeros componentes por que ambos expresan 94.61% de la variabilidad máxima existente en los datos, de los cuales el 76.07% de la variabilidad entre las muestras queda expresada en el primer componente principal y el 18.54% en el segundo componente principal. Además facilita visualizar la distribución de las muestras en un plano bidimensional.

Como puede observarse en la Figura 16, a cada tipo de miel se le asignó un color siguiendo la misma regla de la técnica quimiométrica anterior. Hay cuatro grupos perfectamente bien separados; y uno de ellos se presenta más disperso. El grupo de muestras de miel de cítrico presenta un agrupamiento compacto, su variabilidad esta distribuida a lo largo del componente principal dos (PC2); para la miel de coco, las muestras se distribuyen igual en ambos componentes principales. En el caso de la miel multiflora de Campeche se observa más separada del resto de las muestras de los otros grupos, su distribución es ligeramente mayor en el componente principal dos. Y al igual en la técnica de agrupamientos jerárquicos, la miel multiflora de Tabasco presenta una mayor dispersión de los datos a lo largo del primer componente principal. Sin embargo, aún con la dispersión tan grande es posible trazar una frontera con respecto a los otros grupos de muestras y lograr la diferenciación de los cuatro grupos.

5.1.8. Caras de Chernoff

Una técnica utilizada en este proyecto, que se presenta como una propuesta quimiométrica es la de la representación gráfica multivariada y multidimensional de las caras de Chernoff; técnica que consiste en representar cada variable como un atributo perteneciente a una cara. El poder de las caras de Chernoff es su alta condensación de datos y lo atractivo que resulta la forma de representación de las variables. Scott (1992) encontró que la revisión repetitiva de grandes cuadros de datos es tediosa, sin embargo, las caras de Chernoff pueden mejorar

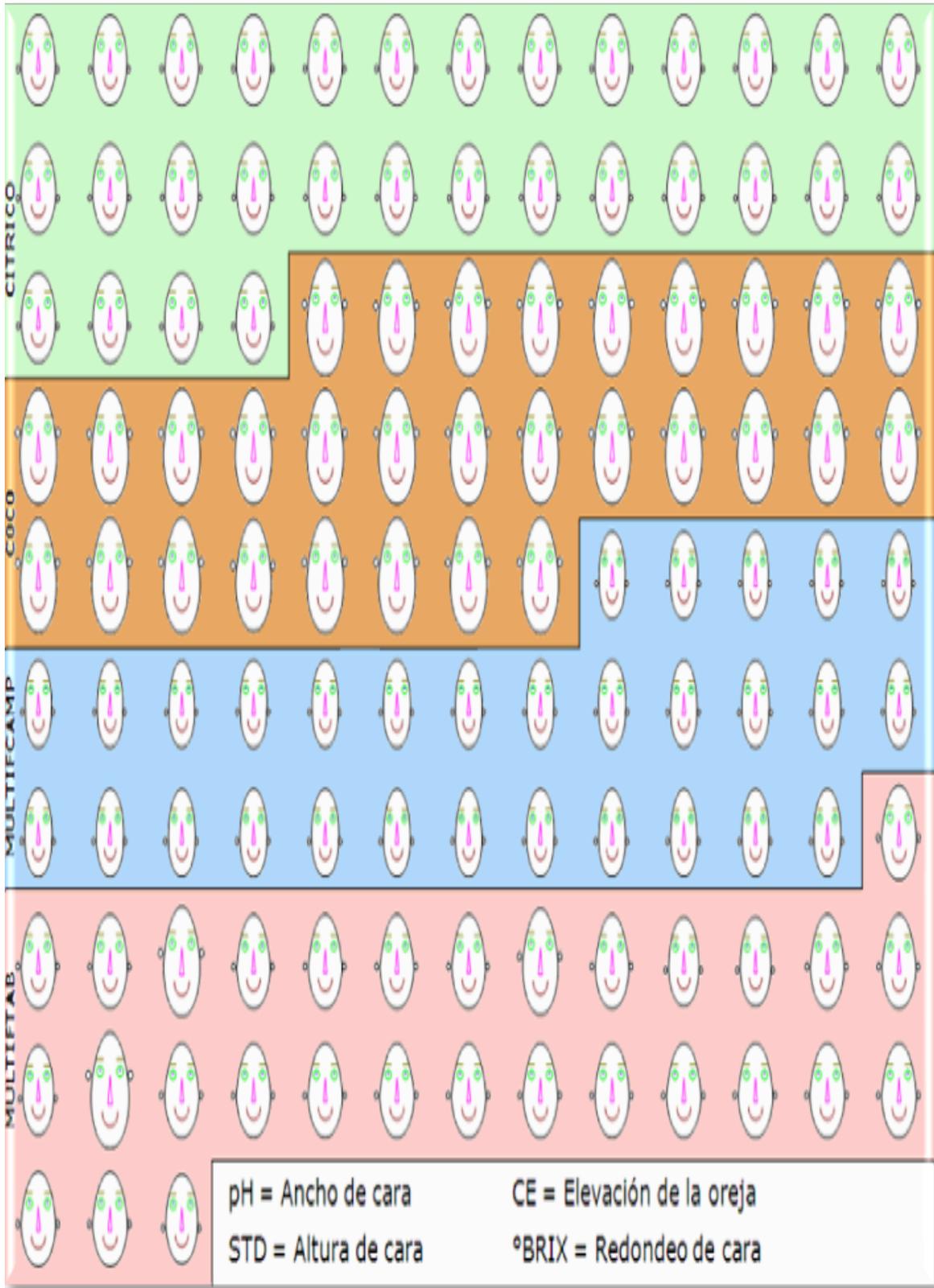


Figura 17. Representación en las caras de Chernoff, de los análisis físico-químicos de las muestras de miel.

significativamente la asimilación de datos. Un inconveniente que enfrentan las caras de Chernoff es que la asignación subjetiva de las expresiones faciales a las variables, afecta sobre la forma de cara. Chernoff y Rizvi (1975) encontraron que la clasificación de dos caras como "bastante similar" está fuertemente influenciada por la asignación de variables a características específicas. Así al aplicar la técnica a los datos obtenidos de los análisis físico-químicos, mediante el programa STATISTICA (Data analysis software system), version 6, StatSoft, Inc. (2003). El resultado es el mostrado a continuación en la Figura 17.

La Figura 17 es la representación gráfica de los datos de las muestras de miel analizadas, cabe señalar que en la figura se presentan las muestras en orden consecutivo, es decir de la muestra 1 a la n , en el caso particular de la muestras analizadas de la muestra 1 a la 120, de izquierda a derecha. Esta forma de representación, permite observar claramente la similitud entre una muestra y la siguiente; siendo así que en el estudio se encontró que las variables determinadas en conjunto caracterizan a cada grupo de muestras. Por ejemplo, en el caso de las muestras de miel de cítrico presentan mucha similaridad pues al analizar las caras generadas estas son muy homogéneas, pero al mismo tiempo presenta diferencias con el siguiente grupo, de muestras de coco, pues estas últimas son diferentes en altura y redondeo de la cara y en la elevación de las orejas.

En el lote de miel de cítricos la altura es menor, es decir, que los valores de STD son menores con respecto a los presentados por el lote de coco, en el caso de la elevación de las orejas de igual manera el lote de cítrico presenta valores menores que el lote de coco, pero con respecto al redondeo al ser las primeras más redondas nos indica que el lote de cítrico presenta valores superiores que el lote de coco, todas estas características en conjunto nos permiten realizar la diferenciación entre estos dos primeros lotes de muestras.

Al compararse el lote de coco y el lote multiflora de Campeche se observaron diferencias en cuanto al ancho, altura, redondeo de la cara y elevación de las orejas, esto es, en el caso del lote de coco esté, al ser más anchas las caras nos indica que los valores de pH son mayores que los presentados por el lote multiflora de Campeche, para la altura podemos decir que los valores del lote de coco son mayores en contenido de STD que los presentados por el lote multiflora por ser más grandes que las caras de estas últimas, en el caso del lote de coco, tiene valores mayores de CE ya que presenta mayor elevación de las orejas que el lote de miel multiflora, por lo que se puede decir que existen diferencias entre los lotes de coco y multiflora.

Por último, en la comparación realizada entre los lotes de miel multiflora de Campeche y multiflora de Tabasco encontramos que en las caras se observan diferencias en el ancho, altura y redondeo de la cara, esto es, en el caso del ancho de la cara al ser más anchas las de la miel de Tabasco nos indica que estas presentan un valor mayor de pH que las de Campeche. Con respecto a la altura de cara se observa que la miel de Tabasco presenta valores mayores en el contenido de STD que las de Campeche; en cuanto al redondeo de la cara, la miel de Tabasco tienen valores mayores en °Brix que la miel de Campeche, mientras que en el caso de la elevación de la oreja no se aprecia mucha diferencia entre ambos lotes, es decir, que la diferencia no es muy pronunciada entre ambos lotes. Sin embargo, estas características físico-químicas permiten observar que existen diferencias entre ambos lotes.

Es importante destacar, como se ha visto en técnicas anteriores, que este último grupo (multiflora de Tabasco) es el más heterogéneo ya que en las caras se pueden apreciar variaciones en las formas.

Con los resultados obtenidos del análisis físico-químico, podemos indicar que se identifican cuatro grupos de mieles perfectamente diferenciados cuando se les

realizaron los análisis quimiométricos. Lo cual resultará de importancia para comprobar la capacidad de la tecnología NIRS para diferenciar mieles en función de su origen botánico o geográfico.

Con los resultados obtenidos en las representaciones gráficas podemos indicar que para la interpretación de los análisis físico-químicos, las que mejores resultados se obtienen para visualizar la diferenciación entre miel de diferente origen botánico o geográfico es la técnica de las caras de Chernoff, análisis de componentes principales, y como una tercera alternativa pero tomando sus reservas el análisis de agrupamiento jerárquico. Aunque cualesquiera de las técnicas aplicadas son factibles, el análisis de agrupamientos jerárquico esta siendo desplazados por el análisis de componentes principales, con lo cual cada vez es menor su uso.

5.2. Análisis NIRS

En el caso de los datos obtenidos en las lecturas de las 120 muestras de miel y dado que el equipo genera un dato por cada dos longitudes de onda y que éstas se presentan en números pares desde los 400 hasta los 2500 nm, se genera un total de 1050 datos o variables por muestra; obtenemos una matriz de datos generada conformada por 126,000 datos. Los 1050 datos al ser graficados representan la huella espectral característica de cada muestra como se muestra en la Figura 18. Los colores fueron empleados para facilitar la diferenciación entre lotes de muestras en donde el café representa las muestras de miel de coco, el verde miel de cítrico, rojo miel multiflora de Tabasco, azul miel Multiflora de Campeche.

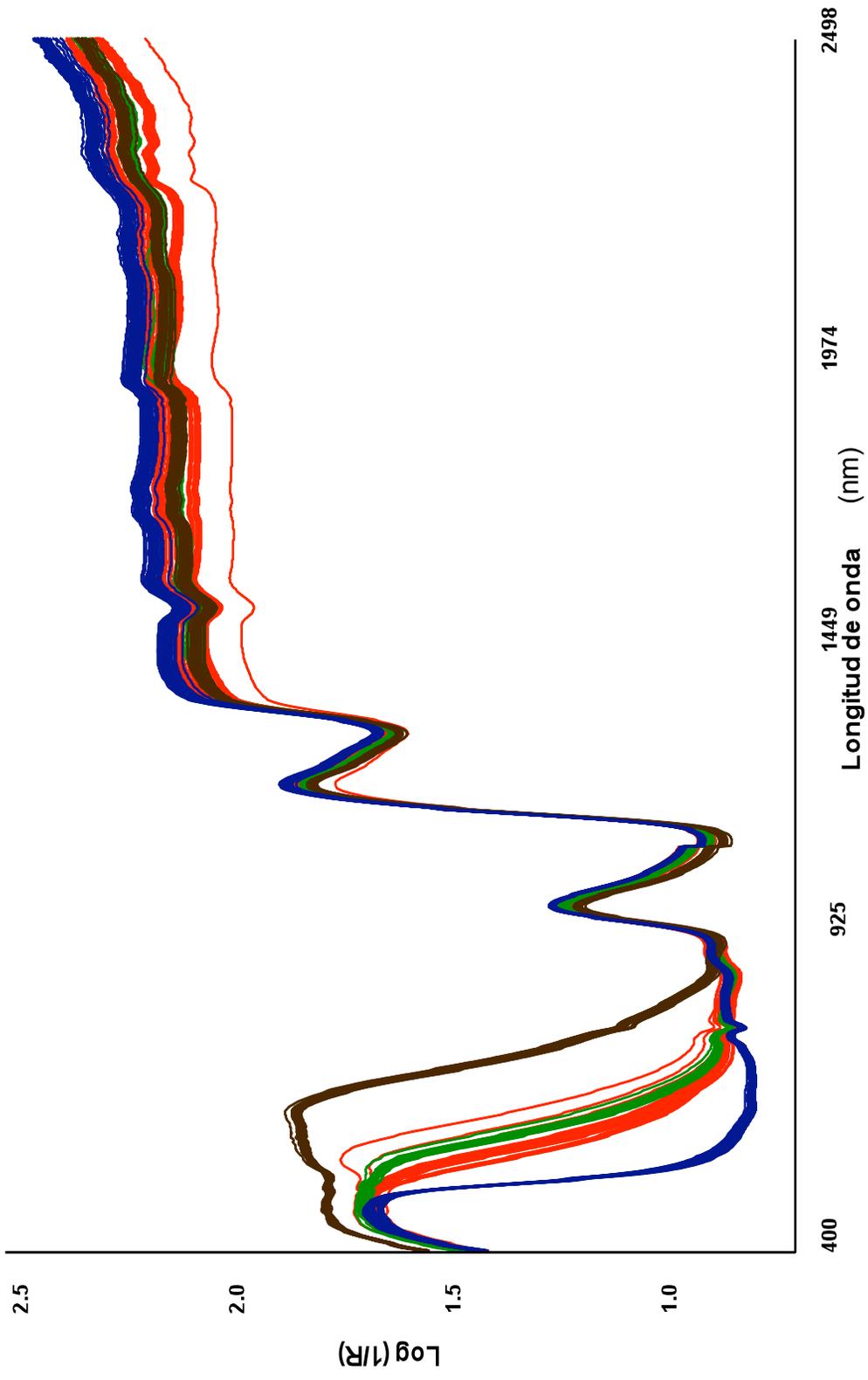


Figura 18. Espectros NIRS de las muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).

Como se puede observar en los espectros existen ciertas regiones que presentan más diferenciación que otras para cada grupo de muestras analizadas. Así mismo el espectro fue construido cubriendo tres regiones, UV-Vis (400 – 700 nm), la región de transición UV-Vis/NIR (702 – 1098 nm) y la región NIR (1100 – 2500 nm).

Regiones que pueden apreciarse si marcamos las tres regiones en diferentes colores (Figura 19).

En el color amarillo se ubica la primera, la región UV-Vis (400 – 700 nm); región en la cual, cuando se hace incidir energía lumínica esta es absorbida y se refleja como vibración electrónica.

La segunda región, en color azul corresponde a la transición entre UV-Vis y NIR (702 – 1098 nm), es en esta región en donde se realiza la identificación de algunos componentes presentes en las muestras, tales como almidón o celulosa que pueden ser encontrados de 858 a 868 nm, 894 a 906 nm y 916 a 918 nm, proteína que puede ser encontrada de 870 a 874 nm, 908 a 910 nm y 1048 a 1052 nm, solo almidón de los 876 a 884 nm, azúcar de 886 a 888 nm y 1004 a 1006 nm, aceites en los 890 a 892 nm, 912 a 914nm, 930 a 932nm, 962 a 968 nm y 998 a 1002 nm, agua de los 934 a 960 nm, 984 a 996 nm y 1008 a 1012 nm, también puede encontrarse aceite o proteína de los 1014 a 1032 nm.

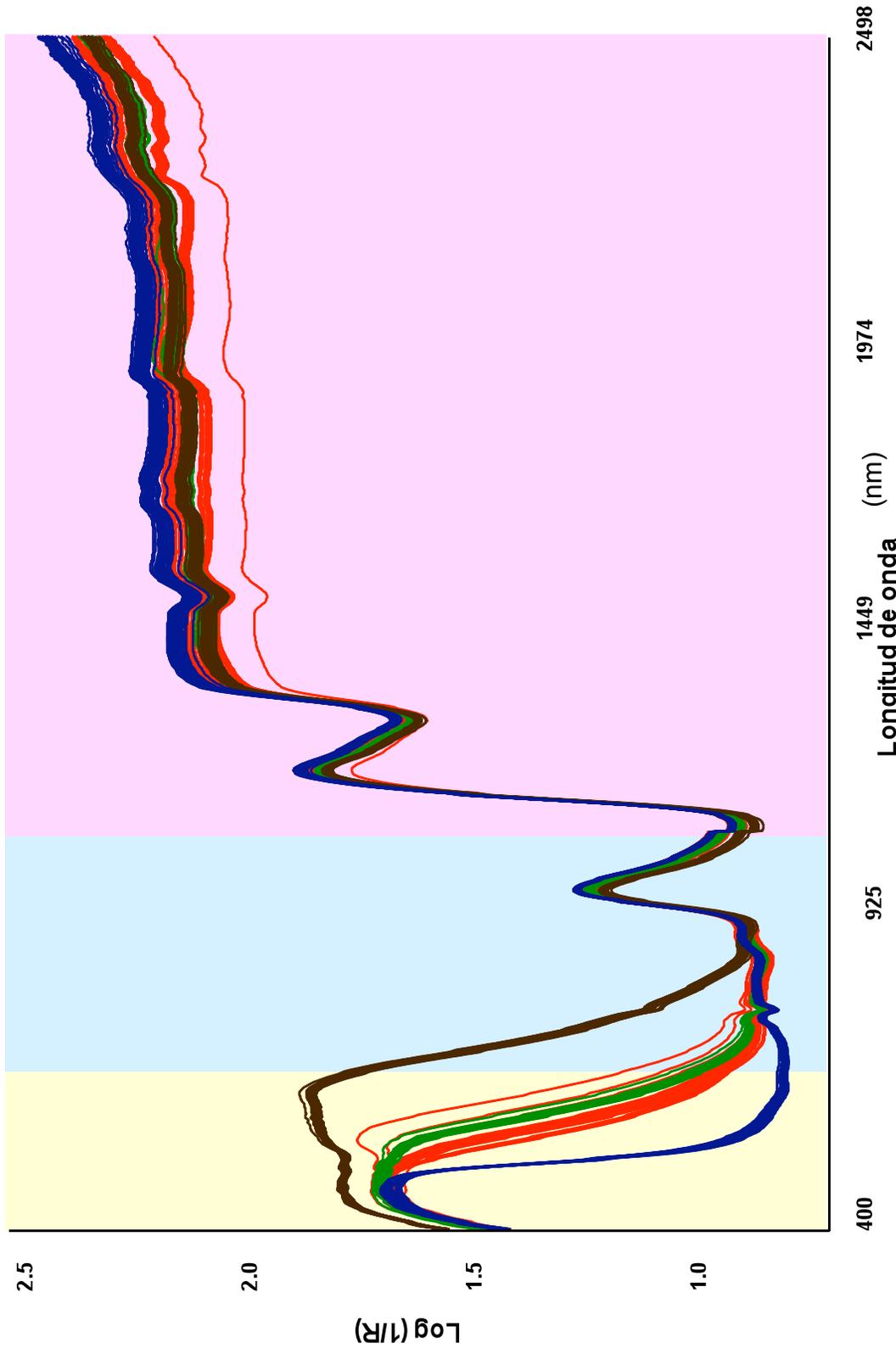


Figura 19. Regiones en espectros generados por cada una de las muestras de miel.

La tercera región, en lila corresponde a la región espectral NIR (1100 – 2500 nm) en donde se presenta únicamente absorción de energía reflejada en estiramiento y doblez de las moléculas; pero independientemente de realizar la identificación de los componentes es precisamente en la primera y en parte de la segunda región en donde los espectros obtenidos en las lecturas de cada una de las muestras presentan la mayor diferencia. Por ejemplo, en la Figura 19 podemos observar que las mieles que presentan mayor vibración electrónica son las mieles de coco, seguida por las mieles de cítrico, y las mieles multiflora de Tabasco, las cuales están más próximas entre si, y por ultimo las mieles multiflora de Campeche. Sin embargo, podemos ver que el comportamiento cambia al observar la abundancia de componentes pues en este caso se observa que la miel multiflora de Campeche es la que presenta mayor abundancia, seguida por un traslape entre las mieles de cítrico y multiflora de Tabasco y por ultimo se observa en la miel de coco, una diferencia más pronunciada sobre todo en la parte cercana a las región de transición entre la región UV-Vis y NIR.

En la región NIR se observa que la miel multiflora de Campeche es la que presenta mayor absorción de energía seguida por muestras de miel multiflora de Tabasco y un traslape de muestras de mieles de cítrico y coco y por último encontramos otras muestras de miel multiflora de Tabasco.

5.2.1. Análisis de agrupamiento jerárquico (HCA)

Con los datos generados con el espectrómetro de infrarrojo cercano se creó una base de datos en WINISI y después a Excel (programa puente entre NIRS y Pirouette), los cuales fueron transferidos al programa quimiométrico Pirouette. Con ellos se realizó, en una primera exploración, un análisis exploratorio mediante la técnica de agrupamiento jerárquico con el cual se obtuvo un dendrograma (Figura 20), utilizando como medio de prueba la distancia Euclídea y el método de unión simple, este nos permitió ver el comportamiento y distribución de los datos, así

como las similitudes y disimilitudes entre muestras, cabe señalar que al igual que en los análisis físico-químicos para poder observar mejor esta distribución a cada tipo de miel se le asignó un color; se utilizó para ello, el patrón anterior, es decir, azul para la miel multiflora de Campeche, verde para la miel de cítrico, café para la miel de coco y rojo para la miel multiflora de Tabasco.

En la Figura 20 se observan la formación de cuatro grandes grupos de muestras de miel bien diferenciados: el de miel multiflora de Campeche, el de las mieles de coco, el de las mieles de cítricos y la miel multiflora de Tabasco; cuyos índices de similitud son: 0.821, 0.837, 0.852 y 0.748, respectivamente.

En el caso de la miel multiflora de Campeche se observan tres subgrupos de muestras los cuales presentan un índice de similitud de 0.821. Con la información que se cuenta no es posible identificar el origen de estos subgrupos, solo las huellas espectrales del NIR. Aun con la formación de estos subgrupos, la miel multiflora de Campeche presentó un índice de similitud alto.

En la miel de coco no es posible observar subgrupos bien definidos lo cual es debido a la probable homogeneidad de las muestras. Por lo que se observa que este tipo de miel presenta un buen agrupamiento, con un índice de similaridad de 0.837 que es un valor superior al obtenido por la miel de multiflora de Campeche.

En el caso de la miel de cítrico presenta el mayor índice de similitud (0.852) de los cuatro grandes grupos. Se puede observar, la tendencia a formar subgrupos pero poco diferenciados.

Por último se puede indicar que en el caso de la miel multiflora procedente del estado de Tabasco al igual que los grupos anteriores y contrario al resultado obtenido con la caracterización físico-química, presenta el agrupamiento esperado. Aunque es evidente que es el lote de muestras que presenta menor

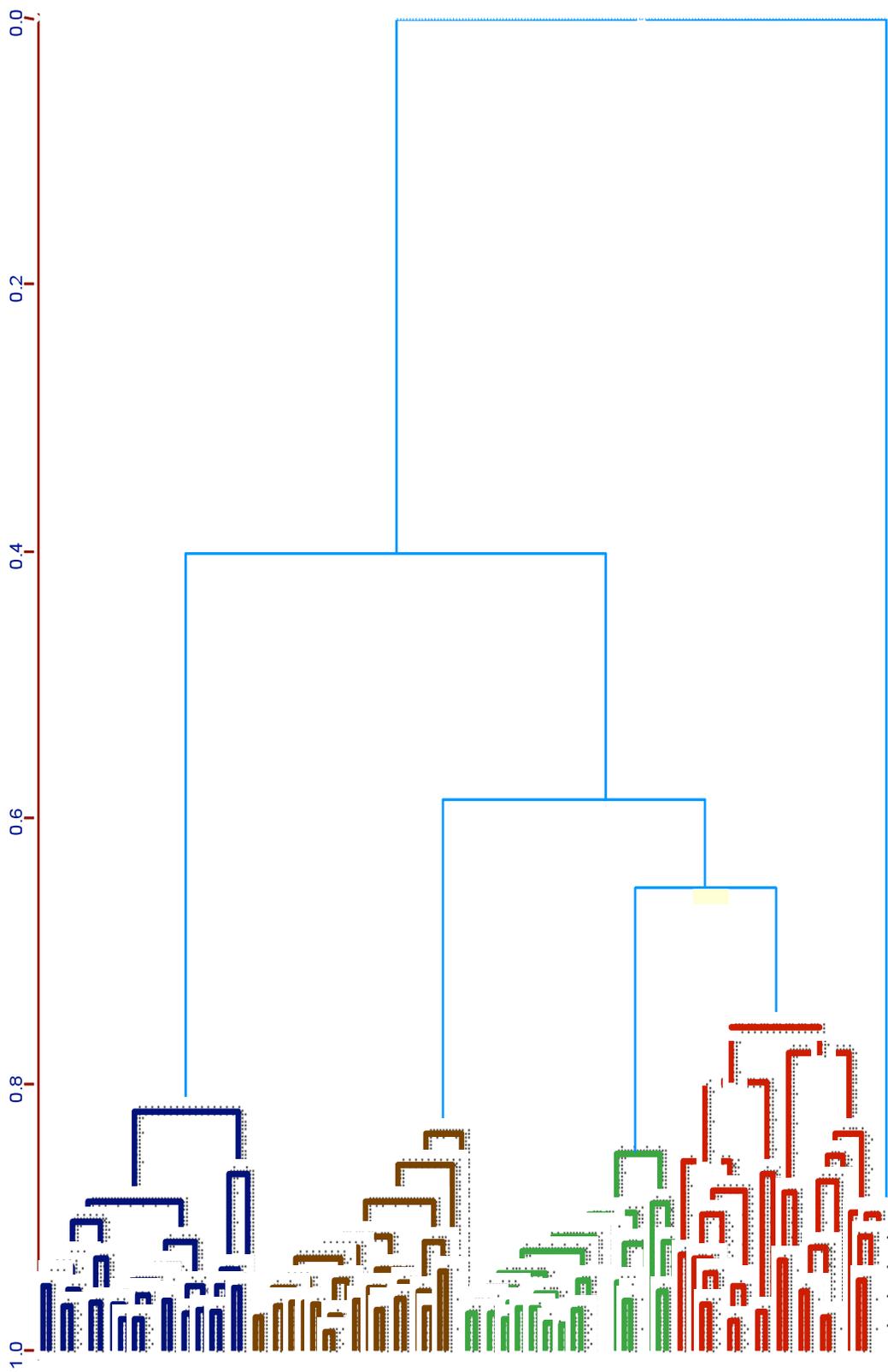


Figura 20. Dendrograma generado con las huellas espectrales NIR de las muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).

homogeneidad; ya que el índice de similitud en este lote es de 0.748, siendo este el valor menor mostrado entre los cuatro grupos de muestras de miel.

Como conclusión de este análisis podemos indicar que se encontraron cuatro grupos de muestras de miel con diferente grado de homogeneidad interna, la cual queda reflejada en los valores de similaridad de cada grupo. La conformación de los cuatro grupos en el dendrograma confirma la calidad de los datos generados con la tecnología NIRS para tipificar muestras de miel de diferente origen botánico o geográfico. Los cuatro grupos de muestras de miel identificados con los análisis físico-químicos son también identificados en el dendrograma que se obtuvo con espectros NIRS.

5.2.2. Análisis de componentes principales (PCA)

Otra de las técnicas de análisis exploratorio aplicada a los datos espectrales NIR, fue el análisis de componentes principales; el cual, nos permitió reducir el número de variables obtenidas al momento de la lectura de las muestras. Es decir, que mediante esta técnica en lugar de trabajar con 1050 datos o variables espectrales, se obtienen nuevas variables, llamadas componentes principales en las cuales con tan solo 10 componentes principales se explica aproximadamente el 100 % de la variabilidad del total de las muestras. Lo que reduce la dimensionalidad de los datos.

Con los componentes principales calculados a partir de las 1050 señales espectrales NIR, se encontró que la máxima variabilidad de las muestras queda expresada en los dos primeros componentes principales, ya que el primer componente principal nos explica el 99.60% de la variabilidad de los datos y el segundo un 0.36% de la restante, es decir, que en conjunto ambos componentes principales nos expresan el 99.96% de la variabilidad total de las muestras de miel. Un valor de la variabilidad total explicada en los dos primeros componentes principales considerado bueno, lo cual indica que entre estos dos componentes

nos presentan casi el total de la variabilidad que inicialmente se encontraba expresada en las 1050 variables originales.

Como se ha visto en la literatura, cada variable utilizada representa una dimensión al momento de realizar un gráfico, por lo tanto, si se grafica con las variables originales se tendrá una figura con n-dimensiones, un gráfico difícil de obtener y al mismo tiempo de interpretar. Sin embargo, mediante el análisis de componentes principales, dado que en los primeros componentes tenemos casi el ciento por ciento del total de la varianza reflejada, el resto de los componentes presentan una aportación mínima, así pues, para visualizar la distribución en el espacio de las muestras, es posible realizar dicho gráfico utilizando únicamente dos primeros componentes, los cuales se representan en la Figura 21.

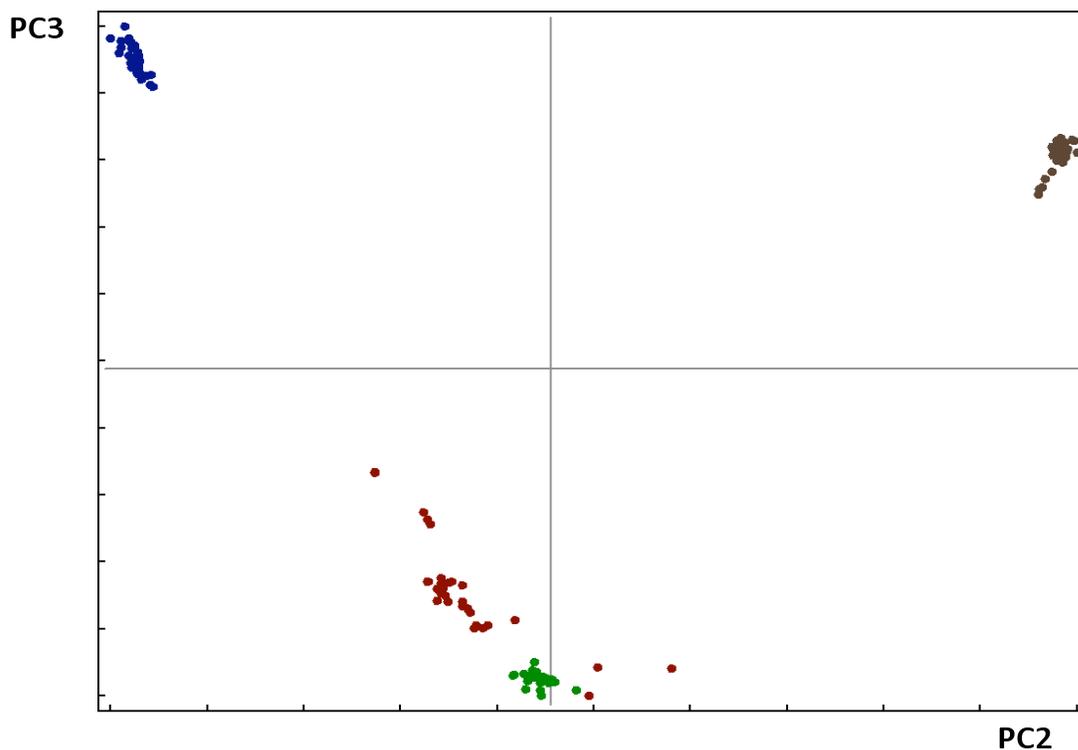


Figura 21. Puntuaciones del conjunto de muestras analizadas en dos componentes principales a partir de las señales espectrales NIRS de las muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).

Cabe señalar que aunque la máxima variabilidad es reflejada en los dos primeros componentes principales, la Figura 21 se generó utilizando el segundo y tercer componente principal. Por ser estos los de mayor poder discriminante de acuerdo con la prueba de Fisher Weight, con lo cual se muestra la siguiente distribución de los objetos en el espacio.

Como se puede observar en la Figura 21, se encuentran cuatro grupos de muestras perfectamente bien definidos, correspondientes a las muestras de miel de coco, cítrico y multiflora de Campeche, y el más disperso es de la multiflora de Tabasco. Así mismo se encontró que las muestras pertenecientes a coco forman un grupo compacto y que sobre este tiene gran influencia el segundo componente, mientras que la miel de cítrico de igual manera forma un grupo compacto pero en este caso la mayor influencia sobre este grupo de muestras la ejerce el tercer componente, y en el caso de la miel multiflora de Campeche se puede observar que forma un grupo compacto, pero que éste se ve fuertemente influenciado por ambos componentes.

En el caso de las muestras de miel multiflora de Tabasco, las muestras se presentan dispersas a lo largo del componente principal uno. Aunque es posible ubicar a todas las muestras en un grupo.

Una vez realizado el análisis exploratorio, HCA y PCA, de las 120 muestras de miel correspondiente a los cuatro grupos de muestras identificados con los análisis físico-químicos. Podemos afirmar que la información generada con la tecnología NIRS es suficiente y útil para diferenciar los mismos grupos identificados con los métodos convencionales.

5.2.3. Modelos de Clasificación y Modelado de Clases.

Una vez que se comprobó la hipótesis planteada en lo que respecta a la capacidad del NIRS para generar información útil y suficiente para tipificar mieles

de diferente origen botánico o geográfico se planteó entonces ver si la información es útil para generar modelos de clasificación. Para ello se propuso establecer los modelos mediante técnicas de reconocimiento de pautas supervisadas como es el análisis discriminante lineal y el modelado independiente de clases.

5.2.3.1. Análisis discriminante lineal (LDA)

Para la aplicación de esta técnica de clasificación al lote se le dió un preprocesado, autoscaling y la primera derivada, y posterior a este se aplicó la técnica de análisis discriminante lineal sobre componentes principales para obtener el modelo de clasificación, debido a la limitante de esta técnica, que no es capaz de trabajar con matrices de datos demasiado grandes. Para lo cual se utilizó el 80% de las muestras como grupo de entrenamiento y el 20% restante como grupo de validación, esta selección se realizó mediante el algoritmo de Kennard-Stone, en el programa KENNARD del paquete V-PARVUS 2008. El modelo fue validado por el método de *cross-validation* con 5 grupos de cancelación, obteniéndose los resultados mostrados en el Cuadro 6.

Cuadro 6. Resultados LDA. Porcentaje de correcta clasificación y predicción para los cuatro tipos de miel.

LDA		MULTIFTAB	CITRICO	COCO	MULTIFCAMP
Habilidad de clasificación		100	100	100	100
Habilidad de Predicción (5 CV)		91.67	100	100	100

Se encontró que la habilidad de clasificación del modelo es del 100%; mientras que la habilidad de predicción queda expresada en un 97.92%. Esto se puede apreciar mejor en la Figura 22 que se muestra a continuación la cual se generó utilizando la diferencia de puntuaciones, de la primera y segunda categoría.

Como se puede observar la técnica quimiométrica aplicada (LDA) es útil ya que se logra la completa separación y diferenciación entre grupos de muestras. La Figura 22 muestra claramente los cuatro grupos diferentes de miel ubicados en sus respectivas clases. De los cuatro grupos se observa que tres de ellos presentan grupos más compactos; siendo el grupo de miel multiflora de Tabasco el más disperso. Sin embargo, esto no influyó en la creación del modelo, ni en su capacidad de predicción; en el Cuadro 6 se indicó que la predicción es la probabilidad de que cada una de las muestras este ubicada en su clase respectiva.

5.2.3.2. Modelado independiente de clases (SIMCA)

La técnica de modelado aplicada en el análisis de los datos generados con el NIRS, fue SIMCA; esta técnica se basa en el análisis de componentes principales, mediante los cuales se construye un modelo independiente para cada una de las cuatro clases, al igual que el caso de la técnica de análisis discriminante lineal, se realizó primero a la matriz de datos espectrales el mismo preprocesado que se aplicó en LDA, con lo cual se pudo observar que las muestras se separan adecuadamente. Pauta que indica la alta probabilidad de éxito en la realización de un modelo de clasificación.

La técnica SIMCA se aplicó como una técnica de modelado sobre componentes principales en datos autoescalados, los modelos fueron creados utilizando el 80% de las muestras como grupo de entrenamiento y el 20% restante como grupo de prueba, el modelo fue validado mediante *cross validation* con 5 grupos de cancelación, con un 95% de confiabilidad, la selección de las muestras para la generación de el grupo de entrenamiento y el grupo de prueba se definió mediante el algoritmo Kennard-Stone, del programa KENNARD del paquete V-PARVUS

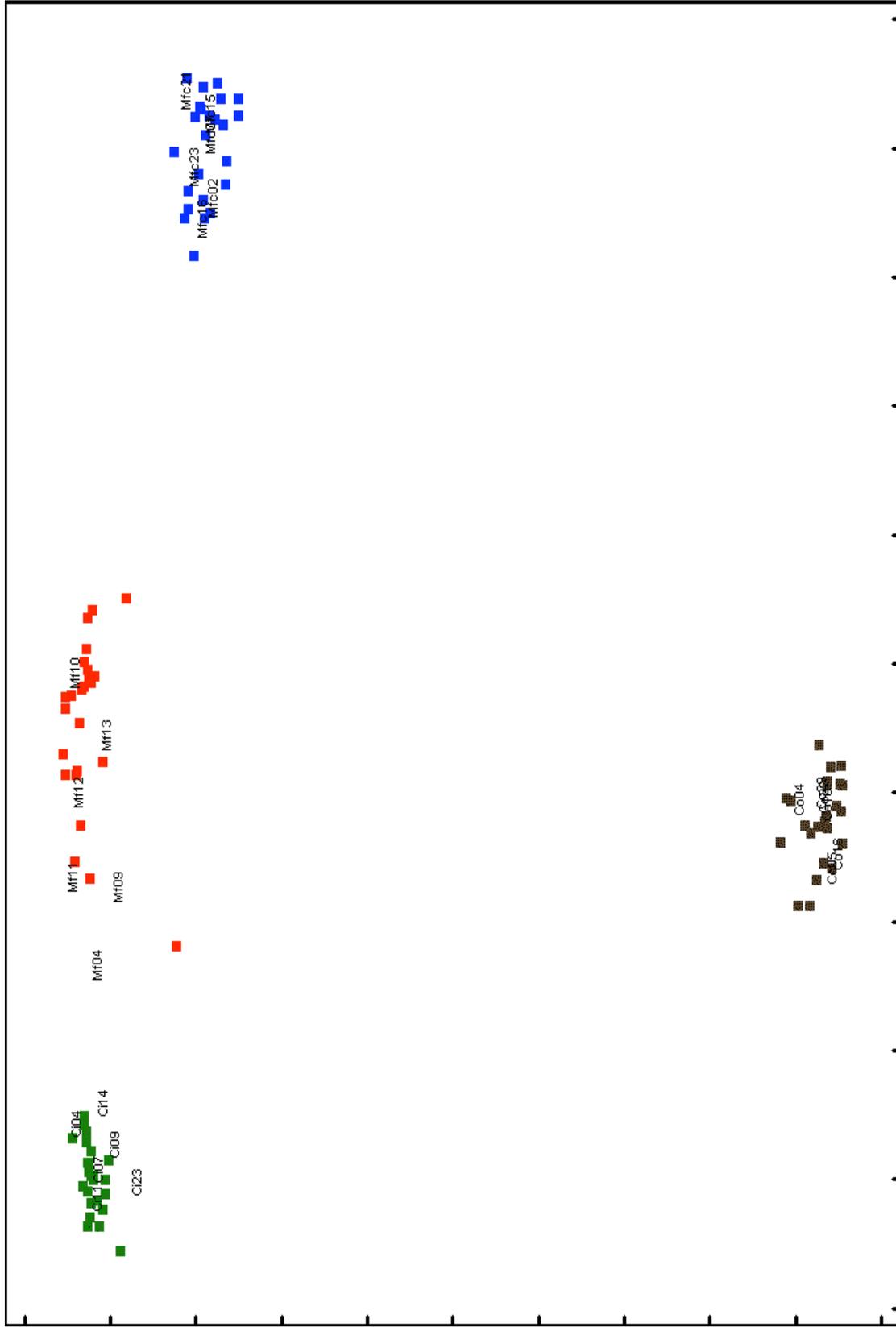


Figura 22. Modelo de clasificación, LDA, obtenido con los espectros NIRS de las muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).

2008 y Pirouette, obteniéndose los resultados mostrados en la Figura 23. Donde mediante la construcción del diagrama de Coomans se realiza el comparativo entre los cuatro modelos diferentes, para sus respectivas clases.

En el diagrama (a) se muestra que al comparar la clase 1, correspondiente a la miel multiflora de Tabasco, con la clase 2, correspondiente a la miel de cítrico; cada una de estas se ubica en su respectiva área de acoplamiento y en el caso de las clases 3 y 4, correspondientes a la miel de coco y multiflora de Campeche, respectivamente, al no presentar semejanza con ninguna de las anteriores son ubicadas en la zona que corresponde a muestras desconocidas, por no pertenecer a las clase 1 y 2.

En el diagrama (b) se representa la comparación entre las clases 1 y 3; cada una de estas se representa en su respectiva área de acoplamiento. En el caso de las mieles 2 y 4 al no presentar semejanza con las dos primeras, son localizadas en el área de no pertenencia a ninguna de las clases comparadas.

En el diagrama (c) se comparan las clases 1 y 4, éstas no presentan semejanza por lo que cada una se ubica en su respectiva área de acoplamiento sin pasar la frontera. Las clases 2 y 3 son ubicadas en el área que nos representa la no semejanza con las clases 1 y 4.

El diagrama (d) esquematiza la comparación entre las clases 2 y 3, y se puede observar que al igual que en los casos anteriores estas presentan disimilitud por lo que cada una de las clases se ubica en su respectiva área de acoplamiento; mientras que las clases 1 y 4, son ubicadas en el área de no pertenencia.

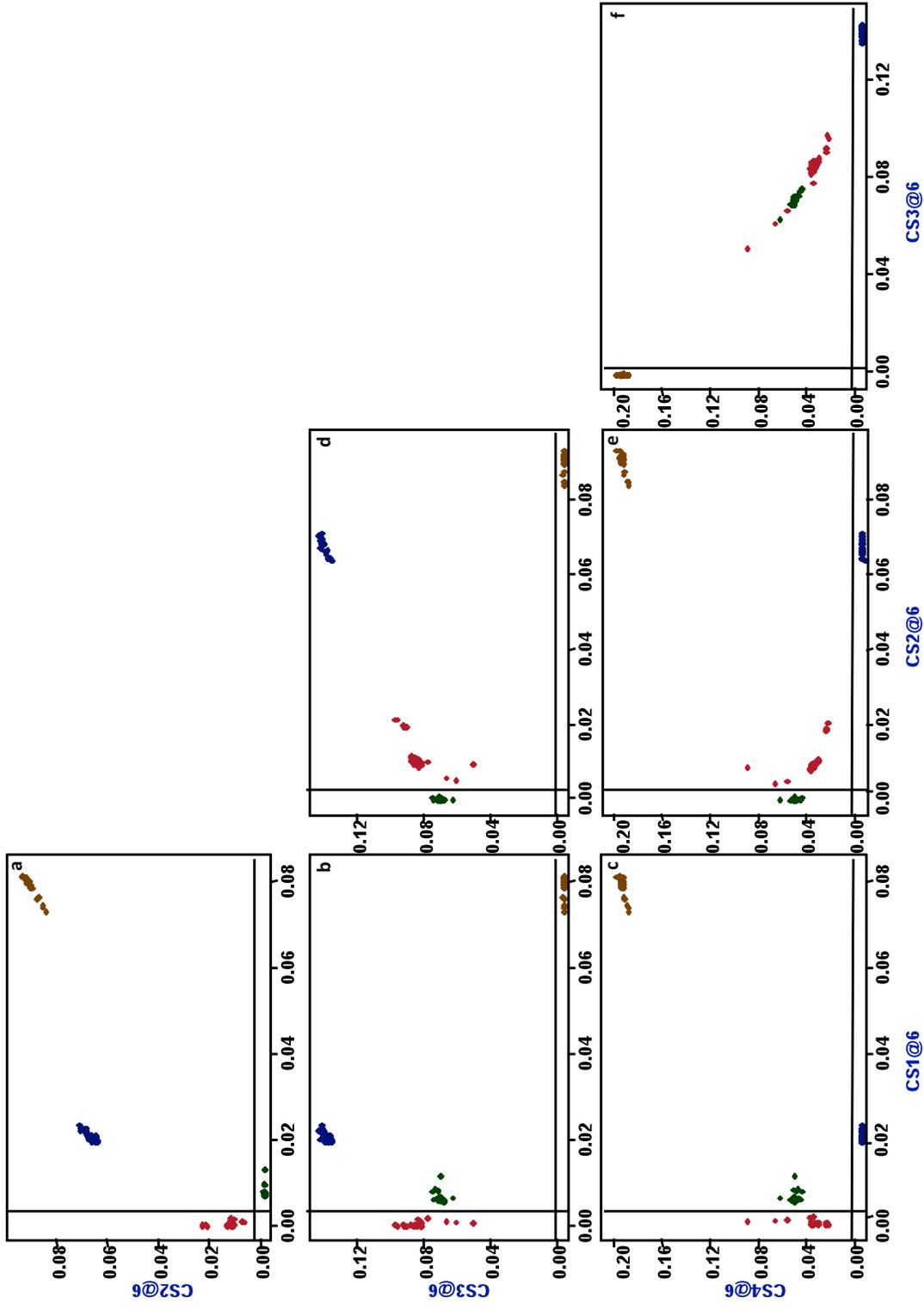


Figura 23. Gráfico del modelado independiente de clases para NIRS a muestras de miel: multiflora Campeche (azul), cítrico (verde), coco (café), multiflora Tabasco (rojo).

En el diagrama (e) se observa la comparación entre las clases 2 y 4, en donde, cada una se ubica en su área de acoplamiento y las clases 1 y 3 ahora son ubicadas en el área de no pertenencia.

Por último, en el diagrama (f) se muestra el comparativo entre clases 3 y 4, y en el se puede observar como cada una de las clases es ubicada en su respectiva área de acoplamiento y en el caso de las clases 1 y 2 se localizan en el área de no pertenencia.

Con los resultados de las comparaciones observadas en la Figura 23, al realizar el comparativo entre clases, ninguna de ellas se mezcla con otra, ni se salen del área de acoplamiento, es decir, que la sensibilidad del modelo es del 100% dado que todas las muestras que componen el modelo son incluidas en el área representativa del modelo, así mismo la capacidad del modelo resultó con un valor del 100%, al rechazar al 100% de las muestras pertenecientes a otras clases para los cuatro grupos de miel.

Por otra parte, al probar el modelo con el grupo de validación, cada una de las muestras desconocidas para los modelos, fue predicha correctamente con lo cual se tienen los resultados que se muestran en el Cuadro 7.

Cuadro 7. Resultados SIMCA. Porcentaje de correcta clasificación y predicción para los cuatro tipos de miel.

LDA	MULTIFTAB	CITRICO	COCO	MULTIFCAMP
Habilidad de clasificación	100	100	100	100
Habilidad de Predicción (5 CV)	100	100	100	100



6

CONCLUSIONES



6. CONCLUSIONES

En el presente trabajo se ha evaluado la capacidad de la espectroscopia de infrarrojo cercano (NIRS), para generar información útil que permita realizar la tipificación de miel de diferente origen botánico o geográfico. Para ello se utilizaron como referencia los resultados obtenidos de la tipificación con los análisis físico-químicos.

A continuación se presentan las conclusiones generales y particulares del trabajo:

El uso de la tecnología NIRS permite la generación de datos el desarrollo de una propuesta metodológica para la tipificación de miel de diferente procedencia botánica o geográfica con la característica de ser rápida, económica y sencilla; cuando se compara con los análisis físico-químicos. Lo que le da mayor relevancia, ya que en la actualidad se busca ahorro en tiempo y dinero, así como reducir la cantidad de muestras en el análisis, reducir el uso de reactivos, entre otros.

Con la investigación realizada se comprobó la hipótesis planteada, y se cumplieron los objetivos establecidos. Con los resultados de los análisis físico-químicos practicados así como la aplicación de técnicas quimiométricas es posible realizar la tipificación de miel de diferente origen botánico o geográfico.

1. En relación con la utilización de las técnicas quimiométricas no supervisadas, análisis exploratorio, se pueden destacar las siguientes conclusiones:
 - Se demostró que en el caso de la técnica de análisis de agrupamiento jerárquico (HCA), para los datos de los análisis físico-químicos, como técnica exploratoria contribuye a detectar la similitud y disimilitud entre muestras y grupos de muestras. Con esta técnica se detectaron los cuatro

grupos de miel que se utilizaron en la investigación. Con cuatro muestras de miel multiflora de Tabasco ubicadas en otros grupos.

- Cuando se aplicó esta técnica a los datos espectrales NIR; los resultados fueron la formación de cuatro agrupamientos perfectamente diferenciados. Cabe mencionar que aunque esta técnica es adecuada para la interpretación de datos multivariantes, actualmente es una técnica ya poco utilizada en quimiometría.
- Cuando se aplicó el análisis de componentes principales, tanto en los datos obtenidos en los análisis físico-químicos como con las señales generadas con la tecnología NIRS, esta técnica demuestra ser adecuada para visualizar la distribución de los datos en el espacio de los primeros componentes principales. Con lo cual es más fácil poder determinar la técnica de clasificación y modelación para proponer una solución simple, rápida y eficaz en la tipificación de mieles de diferente procedencia botánica o geográfica.
- Se utilizaron las caras de Chernoff como parte de los métodos de reconocimiento de pautas no supervisadas, y que han sido poco utilizadas en quimiometría. Su aplicación fue para interpretar los datos obtenidos de los análisis físico-químicos, esta resultó ser una técnica de fácil visualización, para detectar la similitud o disimilitud, entre muestras y grupos de muestras; con lo cual fue posible lograr la tipificación de muestras de miel de diferente procedencia botánica o geográfica.

2. Con respecto al uso de las técnicas de reconocimiento de pautas supervisadas para clasificación y modelado de clases se concluye que:

- El análisis discriminante lineal como técnica de clasificación, resultó ser una técnica eficiente para la correcta clasificación y predicción de las muestras,

es decir, es una técnica que permitió la generación de funciones discriminantes. En el presente estudio el poder de clasificación fue del 100% y un poder de predicción del 97.92%. Por lo que se concluye que el análisis discriminante lineal es una técnica útil para realizar la tipificación de miel de diferente procedencia botánica o geográfica.

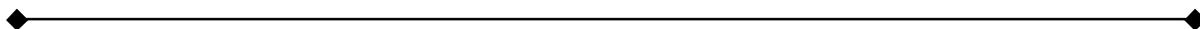
- En cuanto a la técnica de modelado de clases SIMCA, se concluye que en el presente estudio es una herramienta muy útil en la creación de los modelos de clasificación y predicción para cada una de las clases. En los resultados de este estudio se obtuvo un poder de clasificación del 100% y un poder de predicción del 100%. Por lo que resulta útil su aplicación en la tipificación de miel de diferente origen botánico o geográfico.

Con base a los resultados presentados la conclusión general del presente trabajo es que la espectroscopia en el infrarrojo cercano es una herramienta que nos genera la información suficiente y útil que nos permite realizar la tipificación de miel de diferente origen botánico o geográfico, de manera tal que se reduce el tiempo, costo y error del estudio. Lo que la hace una de las herramientas que esta tomando mayor importancia en la actualidad. Así mismo, la espectroscopia en el infrarrojo cercano en conjunto con técnicas quimiométricas apropiadas nos permite visualizar el comportamiento de los datos, estimar una función discriminante, un modelo de clasificación y un modelado de clases que permiten realizar la correcta clasificación y predicción de muestras, el realizar la caracterización y tipificación de miel de diferente origen geográfico o botánico.



7

**LITERATURA
CITADA**



7. LITERATURA CITADA

- Alcantar-González, G., J. D. Etchevers, A. Aguilar. 1992. Los análisis físicos y químicos su aplicación en Agronomía. Centro de Edafología, Colegio de Postgraduados, Montecillos, México.
- Alomar, D. 1998. Fundamentos de la espectroscopia de reflectancia en el infrarrojo cercano como método de análisis de forrajes. *Agro Sur* 26. México.
- Ávila, G., M. Gómez, G. Montenegro y AM. Mújica. 1993. La flora nativa sustentadora de colmenas de *Apis mellifera* en Pichidangui, IV Región de Chile. *Ciencia e Investigación Agraria*. 20 (3), México
- Blanco M. y V. Cerdá. 1988. Quimiometría. Publicaciones de la Universidad Autónoma de Barcelona. Bellaterra. España.
- Bogdanov, S. et al. 1999. Honey Quality and International Regulatory Standards: Review of the Work of the International Honey Commission. *Mitt. Gebiete Lebensm. Hyg.* 90, in press.
- Brereton R., G. 1990. *Chemometrics Applications of Mathematics and Statistics to Laboratory Systems*. Ellis Horwood, Chichester.
- Brereton R., G. 2003. *Chemometrics Data Analysis for the Laboratory and Chemical Plant*. Wiley, Chichester.
- Carvalho A. R., M. Nogal Sánchez, J. Wattoon, and R. G. Brereton. 2005. Comparison of PLS and kinetic models for a second-order reaction as monitored using ultraviolet visible and mid-infrared spectrometry. School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.
- Chernoff, H., & Rizvi, M. H. 1975. Effect on classification error or random permutations of features in representing multivariate data by faces. *Journal of American Statistical Association*, 70, 548-554.
- Codex Alimentarius*. 2001. Codex Standard for Honey. Codex stan 12 (1987), rev. 2.
- Código Alimentario Argentino. Ley 18.284 18/07/69. Artículo 782 - (Res 2256, 16.12.85)

-
- Coomans D., D. L. Massart, and L. Kaufman. 1979. Optimization by statistical linear discriminate analysis in analytical chemistry. *Analytica Chimica Acta*. Polonia.
- Costell E., Duran L. 1981. El análisis sensorial en el control de calidad de los alimentos. Introducción. Instituto de Agroquímica y Tecnología de Alimentos (I.A.T.A.) Rev. A.T.A. 21 (1). Valencia España.
- Cozzolino, D. 1998. Aplicación de la tecnología NIRS para el análisis de calidad de los productos agrícolas. Serie Técnica INIA 97. México.
- Cozzolino, D. 2002. Uso de la espectroscopia de reflectancia en el infrarrojo cercano (NIRS) en el análisis de alimentos para animales. In: *Agrociencia*. Vol. VI No. 2. México
- González Martín I., C. González Pérez, J. Hernández Méndez, and N. Álvarez-García. 2003. Determination of fatty acids in the subcutaneous fase of Iberian breed swine by near infrared spectroscopy (NIRS) with a fibre-optic probe. *Meat Science* 65:713-719.
- Goodacre R. and D. B. Kell. 1996. Correction of mass spectral drift using artificial neural networks. *Analytical Chemistry* 68:271-280.
- Hooper, T. 1976. *Guide to Bees and Honey*. Bland ford Press Ltd. Gran Bretaña.
- Kubelka, P. 1948. New Contributions to the Optics of Intensely Light-scattering Materials. Part I, *J Opt Soc* 38:448-457.
- Macho S. and M. S. Larrechi. 2002. Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *Trends in Analytical Chemistry* 21:799-806.
- Massart D. L., G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, and J. Smeyers-Verbeke. 1997. *Handbook of Chemometrics and Qualimetrics: Part A*. Elsevier, Amsterdam.
- Murray, I. 1986. Near Infrared analysis of forages. In: *Recent advances in animal nutrition*. Edited by W. Haresign and D. J. A. Cole. *Studies in Agricultural and food sciences* Butterworths. Australia.

-
- Muselli, C., H. 2005. Los 6 productos de la colmena. 2ª Edición. La imprenta ya. Argentina.
- Nikolich K., C. Sergides and A. Pittas. 2001. The application of Near Infrared Reflectance Spectroscopy (NIRS) for the quantitative analysis of hydrocortisone in primary materials. J.Serb.Chem.Soc. 66(3). Serbia.
- NMX-F-036-1997. 1997. Alimentos-miel-especificaciones y metodos de prueba. Diario Oficial de la Federación. México.
- Osborne, B. G.; Fearn, T., and Hindle, P. H. 1993. Near Infrared Spectroscopy in Food analysis. Second Edition. Longman scientific and technical.
- Pérès, C. C. Viallon, and J. L. Berdagué. 2002. Curie point pyrolysis-mass spectrometry applied to rapid characterization of cheeses. Journal of Analytical and Applied Pyrolysis 65:161-171.
- Piazza, M.G. 1991. Accorti, M. e Persano Oddo, L.: Electrical conductivity, ash, colour and specific rotatory power in Italian unifloral honeys. Apicoltura 7, 51-63.
- Radovic B. S., R. Goodacre, and E. Anklam. 2001. Contributions of pyrolysis-mass spectrometry (Py-MS) to authenticity testing of honey. Journal of Analytical Applied Pyrolysis 60:79-87.
- Ramis Ramos G., M. C. García Álvarez-Coque. 2001. Quimiometría. Ed. Mc Graw Hill. Síntesis. Madrid. España.
- Salamanca, G. G. 2004. Criterios relativos al análisis sensorial de mieles. Departamento de Química - Facultad de Ciencias Básicas Ibagué, Tolima Colombia.
- Scott, D. W. 1992. Multivariate density estimation: Theory, practice, and visualization. New York. John, Wiley & Sons.
- Tellería, M.C. 2001. El polen de las mieles, un indicador de su procedencia botánica y geográfica. Ciencia Hoy 62: 11 (4-5):63-65.
- Thurston T. J., R. G. Brereton, D. J. Foord, and R. E. A. Escott. 2004. Principal components plots for exploratory investigation of reactions using ultraviolet-



visible spectroscopy: application to the formation of benzophenone phenylhydrazone. *Talanta* 63:757-769.

Tukey, J. 1977. *Exploratory Data Analysis*. Addison-Wesley Publishing Company.

Ulriksen, R., P. 2003. Tesis: Análisis Vegetacional y Palinológico de Especies Dominantes de la Comuna de Litueche, VI Región. Chile.

Wang Z., T. Dean, and B. R. Kowalski. 1995. Additive background correction in multivariate instrument standardization. *Analytical Chemistry* 67:2379-2385

White, J. W; Subers, M.H.; Scheparz, A.J. 1963. The identification of inhibine, the antibacterial factor in honey, as hydrogen peroxide, and its origin in a honey glucose-oxidase system. *Biochimica and Biophysica Acta*.

Yañez, G. M., J. A. Díaz. 1990. Análisis sensorial. Prueba de preferencia-aceptación en frutos de melón con catadores no entrenados (consumidores). Boletín 1. Colegio de Postgraduados. CEICADES. México.

Zhang L., C. W. Small, and M. A. Arnold. 2002. Calibration standardization algorithm for partial least-squares regression: application to the determination of physiological levels of glucose by near-infrared spectroscopy. *Analytical Chemistry* 74:4097-4108.

