UNIVERSIDAD CENTRAL MARTA ABREU DE LAS VILLAS
FACULTAD DE QUÍMICA Y FARMACIA
DEPARTAMENTO DE LICENCIATURA EN QUÍMICA

Tesis de Grado

TÍTULO: CLASIFICACIÓN AUTOMATIZADA DE RONES HAVANA CLUB Y SANTERO MEDIANTE LA ESPECTROFOTOMETRÍA ULTRAVIOLETA-VISIBLE Y LA QUIMIOMETRÍA.

AUTOR: ANDY G. DOMÍNGUEZ RODRÍGUEZ
TUTORES: MSc. JORGE B. DE LA TORRE LÓPEZ
MSc. GEOVANNI J. REYES HERNÁNDEZ
CONSULTANTE: DRA. ISNERI TALAVERA BUSTAMANTE

SANTA CLARA, JUNIO DE 2014

UNIVERSIDAD CENTRAL MARTA ABREU DE LAS VILLAS

FACULTAD DE QUÍMICA Y FARMACIA

DEPARTAMENTO DE LICENCIATURA EN QUÍMICA

Tesis de Grado

TÍTULO: CLASIFICACIÓN AUTOMATIZADA DE RONES HAVANA CLUB Y SANTERO MEDIANTE LA ESPECTROFOTOMETRÍA ULTRAVIOLETA-VISIBLE Y LA QUIMIOMETRÍA.

AUTOR: ANDY G. DOMÍNGUEZ RODRÍGUEZ

TUTORES: MSC. JORGE B. DE LA TORRE LÓPEZ

MSC. GEOVANNI J. REYES HERNÁNDEZ

CONSULTANTE: DRA. ISNERI TALAVERA BUSTAMANTE

SANTA CLARA, JUNIO DE 2014

A la memoria de abuelo Fino,

mi primer maestro.

AGRADECIMIENTOS

Expreso mi eterno agradecimiento y gratitud a todos los que hicieron posible esta investigación:

- A mis padres, mis hermanos, mi amiga Yissel y mis tutores Jorge y Geovanni por su apoyo y sacrificio.
- A los colectivos de Cuadros y Criminalística del MININT de Sancti Spíritus y por su amistad incondicional y contribución permanente, en especial a Emilio, Yoelsis y Raúl.
- A las roneras San José y Paraíso por facilitarnos las muestras.
- A la Dra. Isneri Talavera y a los profesores Aliuska Morales, Luis Bravo, Marta Martínez, Carolina Aguiar por sus valiosos consejos y su ayuda en el momento más oportuno.
- A los colectivos del Área Química del Laboratorio Central de Criminalística y del Centro de Aplicación de Tecnologías de Avanzada por haberme acogido como uno más de ellos y colaborar con este trabajo, especialmente: Santa, Yamilet, Susset, Bustio, René, Thais, Idania, Mayra, Cecilia, Gladys, Indira, Milagros, Katia y Maibel.
- Al colectivo del CNICA de Villa Clara, especialmente a Taimí e Isel.
- A Josefa, Amalia, Sonia, Aracelys, Mixaris, María, Francisco, Cristina y Miriam por su ayuda.

A todos ellos: ¡Muchas gracias!

LISTA DE ABREVIATURAS, SÍMBOLOS Y ACRÓNIMOS

Las abreviaturas símbolos y acrónimos, utilizados en esta tesis para describir algoritmos, técnicas y métodos analíticos, entidades, entre otros, se relacionan a continuación en orden alfabético:

A	Absorbancia
b	Paso óptico o ancho de la cubeta
c	Concentración
C	Constante de Vapnik-Shernovensky
CENATAV	Centro de Aplicaciones de Tecnologías de Avanzada
CV%	Coeficiente de variación (porcentual) o desviación estándar relativa, valor análogo al %RSD, *Relative Standard Deviation*
DAD	Detección por arreglo de diodos, del inglés *diode array detection*
GC-MS	Cromatografía de gases acoplada espectrometría de masas, del inglés *gases chromatography and mass spectrometry*
HCA	Análisis jerárquico de clústeres, del inglés *hierarchical cluster analysis*
HPLC	Cromatografía líquida de alta resolución, del inglés *high performance liquid chromatography*
FID	Detección por ionización de llama, del inglés *flame ionization detection*
GC-MS	Cromatografía de gases acoplada espectrometría de masas, del inglés *gases chromatography and mass spectrometry*
ICP	Plasma inductivamente acoplado, del inglés *inductively coupled plasma*
IR	Infrarrojo
***k*-NN**	Método de los *k* vecinos más cercanos, del inglés *k nearest neighbors*
MININT	En Cuba, Ministerio del Interior
MP	Poder de modelación, del inglés *modeling power*
P	Intensidad del haz de luz emergente
P_0	Intensidad del haz de luz incidente

PCA	Análisis de componentes principales, del inglés *principal components analysis*
PLS-DA	Análisis discriminante sobre la regresión por mínimos cuadrados parciales, de inglés *partials low squares discrimination analysis*
QSAR	Relación cuantitativa estructura-actividad, del inglés *quantitative structure-activity relationship*
RBF	Función básica radial, del inglés *radial basic function*
RMSECV	Error cuadrático medio en validación cruzada, del inglés *root medium square error in crossvalidation*
SIMCA	Modelación blanda e independiente por analogía de clases, del inglés *soft independent modeling by class analogy*
SVM	Máquinas de soporte vectorial, del inglés *support vectors machines*
S_x	Desviación estándar
T	Transmitancia
UV-Vis	Ultravioleta-Visible
v.	Varianza
x	Media aritmética
ε	Coeficiente de extinción molar
λ	Longitud de onda
Φ	Poder de discriminación

RESUMEN

Esta tesis constituye un primer paso en posde lograr la clasificación automática de rones según su marca y tipicidad. Para ello, se desarrollaron y validaron modelos quimiométricos para diferenciarentre las marcas de rones cubanos Havana Club y Santero así como entre sus tipicidades a partir de los espectros UV-Vis.

Se estableció que la modelación de rones oscuros y claros se debe realizar de forma independiente para una mejor discriminación y a una relación de dilución muestra: agua destilada de 180 µL: 2 mL y 1 mL: 2 mL (respectivamente)para el registro de los espectros. Paralelamente se realizó un estudio cualitativo mediante HPLC-DAD comprobándose que los compuestos incorporados durante el añejamiento-acabado del ron determinan las propiedades de absorción UV-Vis.

Para el análisis exploratorio de los datos experimentales se aplicó el PCA y el HCA y se hizo una selección de variables mediante sus poderes de discriminación y modelación estableciéndose idóneo el rango de 215-450 nm para la lectura de los espectros. Se aplicaron los algoritmos de clasificación *k*-NN, SIMCA y SVM, obteniéndose los mejores modelos con el SVM que mostró 100% de sensibilidad en ambos modelos y 100 y 90.91% de selectividad para oscuros y claros respectivamente.

ABSTRACT

This thesis is a first step towards automatic classification of rums by brand and typicality. To accomplish this, chemometrics models to differentiate between brands of Cuban rums Havana Club and Santero and between their typicality from the UV-Vis spectra models was developed and validated.

It was established that the modeling of light and dark rums should be performed independently for better discrimination and a dilution ratio sample: distilled water of 180 µL: 1 mL and 2 mL: 2 mL (respectively) for measuringspectra. In parallel, a qualitative study was performed by HPLC-DAD proving that incorporated compounds during aging rum-finished determine the properties of UV-Vis absorption.

In the exploratory analysis of the experimental data was applied PCA and HCA, and a selection of variables was made by its discrimination and modeling powers. The range of 215-450 nm was established for the reading of the UV-Vis spectra. Classification algorithms *k*-NN, SVM and SIMCA were applied, yielding the best models with SVM that showed 100% sensitivity in both models and 100 and 90.91% selectivity to light and dark respectively.

ÍNDICE

Introducción **1**

Capítulo 1. Visión actualizada del análisis quimiométrico a rones y otras bebidas **4**

1.1. El Ron 4

1.1.2. Elaboración y principales componentes 5

1.2. Autenticación de bebidas alcohólicas: estado del arte 7

1.2.1. Técnicas combinadas 8

1.2.2. Espectrometría de masas 9

1.2.3. Espectrometría de emisión de fluorescencia 10

1.2.4. Espectroscopiainfrarroja 10

1.2.5. Espectrofotometría Ultravioleta-Visible 11

1.3. Espectrofotometría de absorción molecular UV-Vis 15

1.4. Quimiometría 16

1.4.1. Reconocimiento de patrones 17

1.4.2. Sistemas computacionales vinculados al reconocimiento de patrones 20

Capítulo 2. Parte Experimental: Espectrofotometría UV-Vis y Quimiometría **22**

2.1. Materiales y métodos 22

2.1.1. Toma de muestras 22

2.1.2. Medios: instrumental y software 24

2.1.3. Métodos 24

Capítulo 3. Resultados: análisis y discusión **33**

3.1. Registro de los espectros UV-Vis 33

3.1.1. Estudio cualitativo por HPLC 35

3.2. Estudio de la repetibilidad instrumental 35

3.3. Análisis exploratorio de los datos 36

3.3.1. Análisis de componentes principales para rones oscuros 36

3.3.2. Análisis jerárquico de clústeres para rones oscuros 41

3.3.3. Análisis de componentes principales para rones claros 42

3.3.4. Análisis jerárquico de clústeres para rones claros 48

3.3.5. Selección y formación de las clases 49

3.4. Entrenamiento y validación de los modelos de clasificación 50

3.4.1. Modelación para rones oscuros 50

3.4.2. Validación y selección del modelo de rones oscuros 51

3.4.3. Modelación para rones claros 52

3.4.4. Validación del modelo de rones claros 53

Conclusiones 55

Recomendaciones 56

Bibliografía 57

Anexos 62

A.1. Cromatogramas obtenidos por HPLC para el Havana Club Ritual Cubano y el Santero Aguardiente Natural 62

A.2. Tiempos de retención HPLC (bajo las condiciones de trabajo empleadas) para algunos componentes del ron 64

A.3. Características de los espectros UV-Vis de los compuestos orgánicos 66

INTRODUCCIÓN

El bien ganado prestigio y la fama internacional que ostentan los rones cubanos los convierte en blanco de ataques de diversa índole, susceptibles a robos de marca y otras acciones en las que juegan un papel protagónico compañías norteamericanas, apoyadas por la política de bloqueo económico contra Cuba, asistida por el gobierno de los Estados Unidos (Tamayo 2009; Cubadebate 2012).

Sin embargo, internamente también existen obstáculos, el mercadeo de ron adulterado o con insuficiente calidad se observa a menudo, tanto en las vías legales establecidas para su comercialización como en otros mecanismos subyacentes que coexisten al margen de la ley. Este hecho suele ser indicio de varias tipicidades delictivas contra la economía nacional (Ochoa 2003; Rivero and Bertot 2009).

Entre las consecuencias más negativas de las adulteraciones se acentúa la mala imagen del producto transmitida al consumidor, la disminución de las ganancias y más importante aún, las lesiones y daños a la salud de los consumidores (Piola 2011).

Las diferentes formas de adulteración abarcan desde los métodos más simples e inocuos hasta los más complejos y tóxicos. Usualmente se remplaza parte del contenido de la botella por agua, u otra clase de alcohol de menor calidad con concentraciones considerables de sustancias nocivas como el metanol y otros alcoholes superiores (Paz 2005; Johll 2008; Piola 2011).

Para enmascarar el fraude, se aplican procedimientos químicos de baja pericia que dejan residuos tóxicos, proporcionan un sabor desagradable al ron y ponen en peligro la salud de los consumidores. En contraste existen técnicas analíticas para avalar la calidad de las materias primas y del producto (Simunovic 1999; Banguela 2001; Comité Técnico de Normalización 2001; Cenzano 2003; Comité Técnico de Normalización 2007; Johll 2008; Sun 2008; Díaz 2010). No obstante, desde el enfoque metodológico estas son engorrosas y requieren de gran volumen de muestra, imposibilitando su adopción para el trabajo operativo pericial (Johll 2008).

Para cumplir con los requerimientos de la labor operativa se necesitan métodos que con un mínimo gasto de muestra, costo tecnológico y tiempo de ejecución garantice el mínimo de información necesaria para certificar al ron como genuino o reconocer su tipo y marca (Johll 2008).

La quimiometría cumple con estas premisas (rapidez, abaratamiento de costos y miniaturización) que la distinguen como una herramienta capaz de ofrecer los resultados deseados en investigaciones de este tipo (Massart 1999; Tortosa 2011).

En el contexto actual se han desarrollado varias investigaciones relacionadas con la autenticación de las bebidas alcohólicas. Estas se apoyaron en diferentes técnicas analíticas (GC-MS, HPLC, la espectroscopia IR y la espectrofotometría UV-Vis) para la obtención de los datos experimentales, yla quimiometría para el procesamiento y análisis de los datos (Cardoso, Andrade-Sobrinho et al. 2004; Bevin, Fergusson et al. 2006; González-Arjona, López-Pérez et al. 2006; Lachenmeier, Sohnius et al. 2006; Martelo-Vidal, Domínguez-Agis et al. 2013; Markechova, Majek et al. 2014).

Es objetivo del MININT fortalecer el trabajo de las investigaciones criminalísticas y la respuesta pericial en función de un incremento en la efectividad del trabajo de enfrentamiento al delito. La tarea 5R×55* del ministro del interior define una línea de trabajo investigativo para el desarrollo de la quimiometría con fines operativos y en su cumplimiento derivan varios objetivos: papeles y tintas, derivados del petróleo, drogas, suelos y las bebidas alcohólicas.

Sentando bases investigativas en la situación anterior, se plantea el **PROBLEMA CIENTÍFICO**:

> ¿Cómo diferenciar entre los rones Havana Club y Santero así como las tipicidades de cada marca ajustándose a las exigencias de la labor pericial?

Del problema planteado resulta la **HIPÓTESIS**:

* Programa que fomenta el desarrollo investigativo y la modernización tecnológica en el MININT.

Aplicando herramientas quimiométricas de reconocimiento de patrones para el procesamiento de los datos multivariantes procedentes del análisis por espectrofotometría UV-Vis de los rones objeto de estudio es posible su diferenciación por marca y tipicidad.

Para probar la hipótesis de la investigación se proyecta el **OBJETIVO GENERAL**:

Desarrollar modelos para la clasificación automática de los rones objeto de estudio por marcas y tipos a partir de sus espectros UV-Vis y la aplicación de técnicas quimiométricas de reconocimiento de patrones.

A fin de cumplimentar el objetivo general se plantean los **OBJETIVOS ESPECÍFICOS**:

1. Establecer las condiciones del método analítico para un adecuado registro y diferenciación de los espectros UV-Vis.
2. Desarrollar y validar modelos de clasificación usando técnicas de selección de variables, así como PCA y HCA.

CAPÍTULO 1

VISIÓN ACTUALIZADA DEL ANÁLISIS QUIMIOMÉTRICO A RONES Y OTRAS BEBIDAS

El ron se menciona por primera vez en documentos provenientes de Barbados en 1650, muchos coinciden que su nombre proviene de la *vox populi rumbullion*, que significa tumulto (Corominas 1954; Marcano 2010).Se produce desde la antigüedad, pero en las Antillas se inició su producción luego de la introducción de la caña de azúcar en 1493.Ya en los siglos XVII y XVIII era un factor económico de importancia en la región (Marcano 2010). La alta demanda favoreció el desarrollo de este rubro y las exigencias de los consumidores por tener un producto de calidad también fueron creciendo.

En este capítulo se exponen las características esenciales del ron y se contextualizan los retos que enfrenta esta industria: el fenómeno de las adulteraciones y la necesidad de desarrollar técnicas que permitan autenticar el producto.

1.1.EL RON

El ron es una bebida alcohólica obtenida a partir de destilados alcohólicos simples o de la destilación de mostos fermentados de jugos de caña de azúcar, mieles, melazas o sus mezclas. El alcohol se añeja convenientemente en recipientes de roble blanco, de modo que al terminar el proceso se conserven los principios aromáticos a los que el producto debe sus características organolépticas específicas. En su terminación es permitido el uso de caramelo para la corrección del color y carbón activado para la decoloración(Dromi and MERCOSUR 1996; Simunovic 1999; Comité Técnico de Normalización 2001; Díaz 2010).
Los rones cubanos son rones ligeros, producto de un aguardiente ambarino, transparente y cuyas características particulares se deben a la madera de roble blanco americano o europeo, que conforma los barriles o bocoyes carbonizados en su superficie interna y dentro de los cuales es envasado para su añejamiento. También debe sus características a las

proporciones de otras sustancias añadidas, al tiempo de añejamiento, las proporciones de caramelo añadido y al conjunto de sustancias con que es mezclado (Cordoví 1988; Iglesias 2003; Marcano 2010).

1.1.2. ELABORACIÓN Y PRINCIPALES COMPONENTES

El proceso tradicional de elaboración del ron puede resumirse en cinco etapas fundamentales: extracción, fermentación, destilación, añejamiento y procesamiento final. (Huertas 2006; Torre 2010)

De forma general la composición química del ron guarda estrecha relación con el tipo de producto, con las características del proceso de elaboración empleado por cada productor, las materias primas utilizadas, sustancias añadidas y el tiempo de añejamiento.

Fundamentalmente se encuentran en el ron ácidos orgánicos volátiles (ácido acético) y otros ácidos no volátiles, ésteres (acetato de etilo), alcoholes superiores (alcohol amílico) y aldehídos (acetaldehído y furfural). También son representativos el metanol, los polifenoles, los taninos y los compuestos nitrogenados y del azufre, estos últimos en concentraciones muy bajas. Es posible encontrar asimismo acetatos y otros compuestos que de conjunto con el etanol como componente principal forman el grupo de los más de 300componentes identificados en el ron (Banguela 2001).

Los componentes mayoritarios del ron se incorporan al formulado en dos períodos fundamentales: fermentación-destilación y añejamiento-acabado.

Fermentación-destilación

El alcohol A, el superfino, al igual que el aguardiente, son las materias primas fundamentales. Estos son obtenidos por vía fermentativa.

Durante la fermentación aparecen otros congéneres además del etanol. Estos componentes permanecen en la mezcla, aún después de la destilación, pues no es posible separarlos porque presentan similares temperaturas de ebullición. Los congéneres tienen una gran influencia en las reacciones que ocurren durante el añejamiento.

Añejamiento-acabado

Durante el añejamiento, se incorporan al aguardiente algunos de los componentes de la madera, principalmente lignina, taninos, hemicelulosas y lactonas.

La lignina es un constituyente de la pared celular de la madera, a la cual aporta rigidez y fortaleza. Contiene fundamentalmente alcohol coniferílico y alcohol sinapínico que experimentan oxidaciones produciendo al final vainillina, ácido vainíllico y ácido siríngico (Klages 1968).

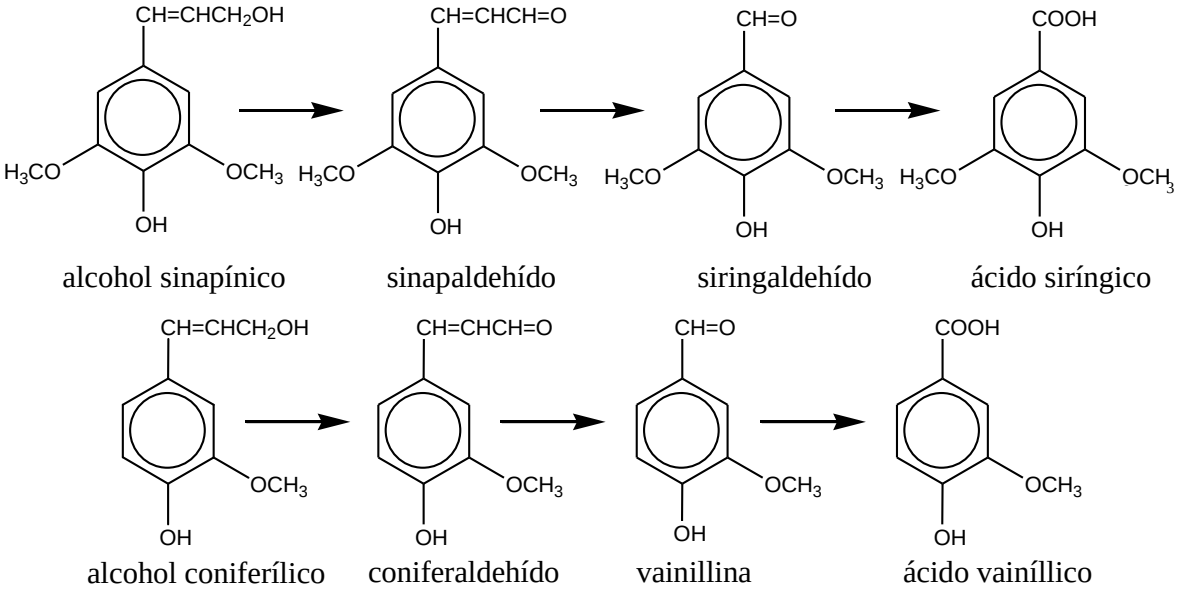

Figura 1-1. Oxidación de los componentes que aporta la hidrólisis de la lignina durante el añejamiento.

Los taninos se encuentran presentes en la madera de roble blanco y le comunican al destilado un sabor astringente. Los polifenoles constituyen aproximadamente un 12% del total de los taninos. El cobre, procedente de la columna de destilación, juega un papel importante en la oxidación de los taninos, este es reducido en el proceso (González 2003).

En las paredes celulares de las plantas existen fibrillas de células densamente empaquetadas, casi cristalinas, estas están unidas por una matriz constituida por tres materiales poliméricos, las hemicelulosas son uno de ellos (González 2003). En los barriles

las hemicelulosas se hidrolizan y forman monosacáridos que aportan dulzor en la bebida (Paturau 1989; Comité Técnico de Normalización 2001).

Las lactonas juegan un papel importante en el aroma del ron. En el ron han sido identificadas la γ-nonalactona, la δ-octalactona y la 1β-metil-γ-octalactona (González 2003).

Ya en el acabado o etapa de procesamiento final son incorporadas otras sustancias como el caramelo para corregir el color. Este al ser preparado con azúcar invertido es rico en glucosa y levulosa. También se adiciona en ocasiones especias y esencias frutales de compleja composición (Pascual 2001; Díaz 2010).

1.2. AUTENTICACIÓN DE BEBIDAS ALCOHÓLICAS: ESTADO DEL ARTE

El acto de adulteración es una acción delictiva que consiste en corromper los alimentos, sustancias o bebidas destinadas al comercio alimentario mediante el empleo de aditivos u otros agentes no autorizados, susceptibles de causar daños a la salud de las personas (Rodríguez and Gallego 1999). Estos hechos fraudulentos son tan antiguos en la sociedad como la historia misma. En respuesta a ello, métodos para legitimar los productos se desarrollaron en paralelo.

Con la creciente conciencia de la seguridad alimentaria y la calidad, los consumidores demandan continuamente la garantía de origen y el contenido de sus alimentos. Además, los fabricantes deben ser capaces de confirmar la autenticidad de los componentes de sus productos con el fin de cumplir con la legislación del gobierno. La protección de los derechos de los consumidores, y la prevención de prácticas fraudulentas o engañosas, así como la adulteración de los alimentos es un desafío importante al que debe hacer frente la industria alimentaria y las fuerzas del orden interior.

En este epígrafe se valorará brevemente cómo y cuánto se avanzó en la autenticación de las bebidas alcohólicas en los últimos años, qué resultados han aportado hasta el momento algunas de las investigaciones desarrolladas, y qué técnicas analíticas son factibles de

aplicar en el contexto de esta investigación. Para un mejor entendimiento se debatirá según las técnicas empleadas de mayor a menor complejidad.

1.2.1. TÉCNICAS COMBINADAS

La combinación de diferentes métodos para la obtención de los datos permite conseguir un amplio número de variables para los modelos de quimiometría.

En 2004 se realizó un estudio comparativo entre la cachaza y el ron valiéndose de los datos analíticos cuantitativos relativos a los componentes orgánicos mayoritarios: achololes (metanol, propanol, isobutanol, isopentanol), acetaldehído, acetato de etilo, ácidos carboxílicos (ácido octanoico, ácido decanoico, y ácido dodecanoico), y polifenoles (ácido protocatéquico, sinapaldehído, siringaldehído, ácido elágico, ácido siríngico, ácido gálico, epícatequina, ácido vaníllico, vainillina, ácido *p*-cumárico, coniferaldehído, alcohol coniferílico, kaempferol y quercetina); además la concentración de iones metálicosde: aluminio, calcio, cobalto, cobre, cromo, hierro, magnesio, manganeso, níquel, sodio y cinc(Cardoso, Andrade-Sobrinho et al. 2004).

Para la determinación de los componentes orgánicos se utilizaron métodos cromatográficos (GC-MS, GC-FID, y HPLC-UV-Vis). La espectrometría de emisión atómica (con ICP) fue empleada en la determinación de los iones metálicos. A continuación los datos analíticos de los compuestos analizados fueron procesados mediante el PCA, HCA, PLS-DA y *k*-NN. Como resultado se logró una muy buena discriminación entre las dos clases de bebidas (Cardoso, Andrade-Sobrinho et al. 2004).

Otro estudio bastante similar, aunque menos profundo, fue realizado en 2006 pero esta ocasión se aplicó a bebidas típicas mexicanas destiladas del agave: tequila, mezcal, sotol y bacanora. La investigación se centró en cuantificar seleccionados constituyentes volátiles mediante cromatografía de gases y algunos aniones por cromatografía iónica, no se realizó el procesamiento quimiométrico de los resultados experimentales (González-Arjona, López-Pérez et al. 2006).

En 2013 se hizo un análisis espectral a los vinos de la denominación de origen Rías Baixas. Los espectros en las zonas IR cercano y UV-Vis combinados con los principios de quimiometría constituyeron los medios empleados para la evaluación. Se utilizó el PCA para explorar y determinar patrones de datos, y la SIMCA y el PLS-DA para la clasificación (Martelo-Vidal, Domínguez-Agis et al. 2013).

Las muestras las componían vinos de las subzonas Condado, Salnés, Rosal y Ribeira do Ulla, todas dela variedad Albariño. Se obtuvieron espectros de transmitancia en el IR cercano y UV-Vis para elaborar los modelos de clasificación. Los resultados demostraron la efectividad de la combinación de espectros IR cercano y UV-Vis conlos métodos quimiométricos consolidándolos así como metodologías rápidas para clasificar subzonas en una denominación de origen, como Rías Baixas. Los modelos obtenidos también fueron capaces de discriminar vinos de otras denominaciones por lo que puede ayudar a la autenticación(Martelo-Vidal, Domínguez-Agis et al. 2013).

Como se ha visto, el uso combinado de disímiles técnicas para la diferenciación entre bebidas alcohólicas y su autenticación ofrece muy buenos resultados, condicionados en gran medida por la variedad de parámetros cuantitativos considerados como criterios de comparación y por la sensibilidad de los métodos empleados. Sin embargo, el sobreprecio y la mayor complejidad de los análisis son factores que van en detrimento de la generalización de este proceder, sobre todo en laboratorios con recursos limitados.

1.2.2. ESPECTROMETRÍA DE MASAS

Un análisis de los congéneres volátiles en cincuenta y dos whiskies comercializados, de los tipos escocés de malta, bourbon e irlandés, se llevó a cabo en 2006 mediante GC-MS después de haber realizado extracción líquido- líquido con diclorometano. Para la discriminación de las diferentes categorías de whisky se aplicaron procedimientos de reconocimiento de patrones. Conjuntamente, el análisis multivariante de los datos incluyó el PLS-DA, el *k*-NN, la SIMCA, y técnicas artificiales de redes neuronales (González-Arjona, López-Pérez et al. 2006).

Para validar las normas de clasificación se consideró el número de falsos positivos y falsos negativos de cada clase asociada al conjunto de predicción. Fueron las redes neuronales artificiales las que condujeron a los mejores resultados por tener características intrínsecas no lineales. El resto de los análisis mostraron también buenos resultados pero con determinados ajustes y limitaciones, sobre todo la SIMCA, atribuyéndose estos resultados a la fuerte influencia de la falta de homogeneidad de la clase (González-Arjona, López-Pérez et al. 2006).

1.2.3. ESPECTROMETRÍA DE EMISIÓN DE FLUORESCENCIA

Para detectar adulteraciones en el brandy se desarrolló un método basado en espectros de excitación-emisión de fluorescencia de la matriz. Los espectros combinados con el análisis factorial paralelo de los datos y la regresión de mínimos cuadrados parciales se utilizaron para determinar el contenido de agua, etanol y metanol en muestras de brandy adulterados (Markechova, Majek et al. 2014).

El espectro de emisión de fluorescencia se midió en el rango de longitud de onda de 510 a 600 nm y el intervalo de 393 a 497 nm fue el establecido para la obtención de los espectros de excitación de fluorescencia (Markechova, Majek et al. 2014).

El modelo creado fue capaz de predecir el agua, el etanol y el nivel de metanol en aguardiente adulterado con bajo error cuadrático medio de los valores de predicción, respectivamente, y los coeficientes de determinación de la predicción entre la referencia y los valores predichos mostraron muy buenos valores (Markechova, Majek et al. 2014).

1.2.4. ESPECTROSCOPIAINFRARROJA

Un método rápido y sencillo fue desarrollado para la confirmación de autenticidad del vino tinto durante su transportación y procesamiento; denominado, sistema dactiloscópico del vino. Cuando el vino se transporta de un lugar a otro, se toma una muestra y se obtiene un espectro infrarrojo. Un método de índice de similitud fue utilizado como herramienta para clasificar muestras de vino sobre la base de sus datos espectrales (Bevin, Fergusson et al. 2006).

Los resultados mostraron que se obtuvieron altas tasas de clasificación cuando se analizaron muestras de vino escaneados en diferentes instrumentos. El índice de similitud se demostró para proporcionar una medición aceptable para la autenticación de la integridad del vino tinto durante el transporte, lográndose una correcta clasificación en el 98% de los vinos sometidos a estudio (Bevin, Fergusson et al. 2006).

En cinco de los seis conjuntos de datos de la bodega, el índice de similitud fue clasificado correctamente. También se observó que menos del 1% de los vinos fueron clasificados erróneamente entre las diferentes bodegas investigadas (Bevin, Fergusson et al. 2006).

No obstante a estos resultados, se concluyó necesario realizar más estudios para probar la aplicabilidad del índice de similitud en una situación comercial y para evaluar su potencial como herramienta de control de calidad rápido para su uso rutinario para autenticar las muestras de vino durante el transporte (Bevin, Fergusson et al. 2006).

En otra ocasión se recurrió a los espectros del infrarrojo medio para el análisis y discriminación de coñac y otras bebidas destiladas: armañacs, whisky, brandy, borbones, rones, y productos falsificados. Se encontraron fuertes correlaciones entre los espectros del extracto seco, los espectros del extracto polifenólico seco, y la concentración de polifenoles totales de las muestras, especialmente del coñac (Picque, Lieben et al. 2006).

Al realizar el análisis de componentes principales aplicado a los datos espectrales se hizo evidente la importancia de los datos de extracto seco cuando se trata de hacer una distinción entre el coñac y armañac, whiskies, whisky americano, y rones, y de la concentración de polifenoles cuando se trata de separar el coñac, el brandy, y los productos falsificados (Picque, Lieben et al. 2006).

1.2.5. ESPECTROFOTOMETRÍA ULTRAVIOLETA-VISIBLE

Los espectros UV-Vis, comparados con las técnicas antes abordadas, no aportan mucha información o mejor dicho, no son muy explícitos. Muchos se muestran escépticos de obtener buenos resultados mediante esta técnica. Sin embargo, la práctica ha demostrado que pueden ser muy valiosos y ofrecer excelentes resultados en la clasificación y

certificación de las bebidas alcohólicas. A continuación se analizarán los resultados de dos investigaciones en las que los espectros UV-Vis son piedra angular.

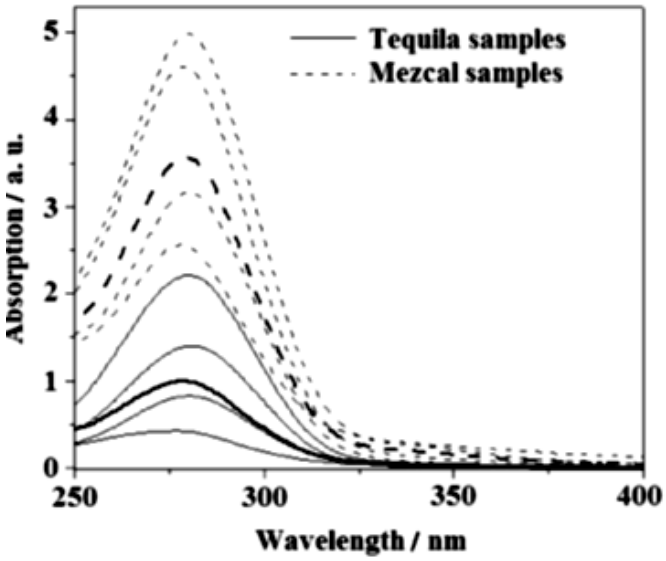

Figura 1-2. Espectros UV-Vis de varios tequilas y mezcales (en negrita se señalan los espectros medios) (Muñoz-Muñoz, Pichardo-Molina et al. 2010).

La primera se desarrolló en 2010 con el fin de lograr la identificación y cuantificación de compuestos furánicos en el tequila y el mezcal usando Espectroscopia y métodos quimiométricos. En efecto, el tequila y el mezcal fueron fácilmente discriminados mediante el uso de la radiación UV-Vis y las técnicas quimiométricas (Muñoz-Muñoz, Pichardo-Molina et al. 2010).

Los espectros se registraron en el intervalo de 250 a 400 nm y la característica principal es una amplia banda centrada alrededor de 280 nm, cuya amplitud cambia según el tipo de bebida y marca. Sin embargo, se señalaron diferencias adicionales a través de un análisis multivariado que hace una diferenciación justa entre las bebidas (Muñoz-Muñoz, Pichardo-Molina et al. 2010).

Mediante el uso de RP-HPLC con detector UV se determinaron los compuestos que originan la banda de absorción. Por lo tanto, el espectro de cada bebida en el rango de longitud de onda de 250 a 330 nm, aproximadamente, es una mezcla particular de derivados

furánicos. Además, los espectros de absorción resultaron consecuencia de mezclas de derivados furánicos (Muñoz-Muñoz, Pichardo-Molina et al. 2010).

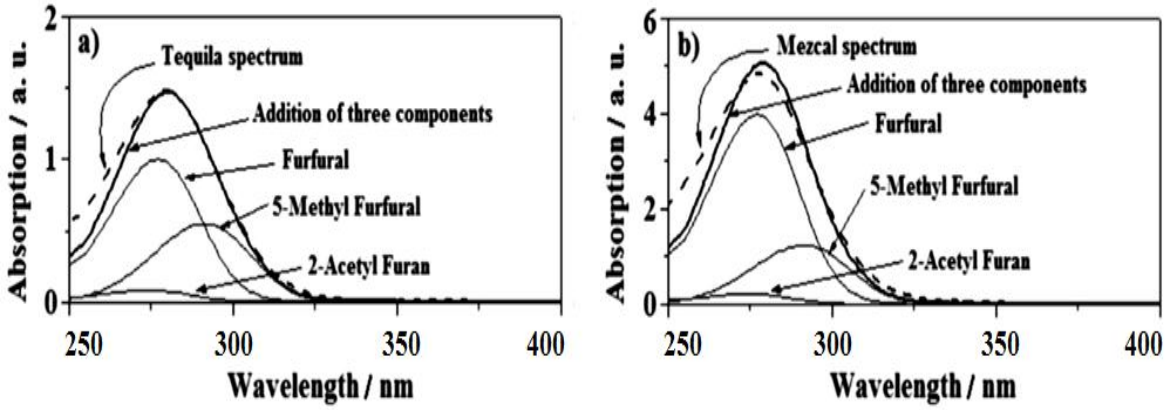

Figura 1-3. Comparación de espectros UV-Vis obtenidos para cada componente y sus mezclas en las proporciones en que se hallan en: (a) el tequila, (b) el mezcal(Muñoz-Muñoz, Pichardo-Molina et al. 2010).

En resumen, se midió con HPLC la concentración y proporción de los derivados furánicos en los tequilas blancos tequilas y mezcales. Para el caso de tequilas blancos, todos estos compuestos exhibieron las concentraciones más altas en el agave tequilera que en tequilas mixtos de otros agaves. Mediante la comparación de tequila blanco y mezcales, la concentración promedio de contenido furánico fue mayor en mezcales (Muñoz-Muñoz, Pichardo-Molina et al. 2010).

Los resultados obtenidos por HPLC corroboraron los obtenidos por aplicación de los métodos quimiométricos a los espectros UV-Vis de estas bebidas. Asimismo, el estudio estadístico desarrollado no detectaron diferencias significativas entre la determinación por HPLC y la realizada por mediante el modelo de regresión lineal de mínimos cuadrados, resultado que avaló al modelo como un método rápido, confiable y eficaz para analizar estas bebidas (Muñoz-Muñoz, Pichardo-Molina et al. 2010).

La segunda investigación publicada en el 2011; propone un analizador automático de flujo en serie para la detección de la adulteración de whisky, brandy, aguardiente, ron y vodka mediante espectrofotometría UV-Vis. Este analizador fue usado para diluir las muestras no

adulteradas y simular la adulteración con adición de diferentes proporciones de agua, etanol o metanol (Nascimento, Araújo et al. 2011).

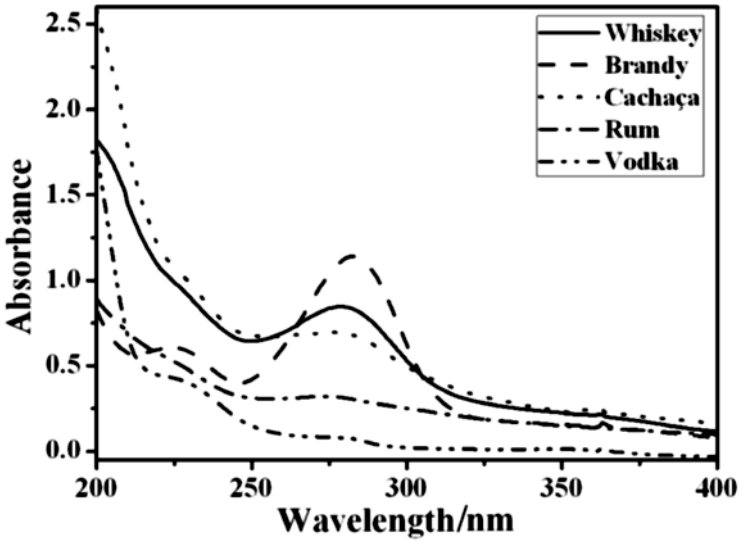

Figura 1-4. Espectros medios de la absorbancia UV-Vis de varias bebidas(Nascimento, Araújo et al. 2011).

Se recurrió al análisis de clase por modelación suave independiente para lograr modelos de clasificación usando espectros en la región de 235 a 355 nm de muestras adulteradas y no adulteradas. Solo se consideró esta región del espectro porque las longitudes de onda mayores "*no presentaron información espectral relevante*" y las menores "*presentaron absorbancias muy altas, cercanas a la saturación*" (Nascimento, Araújo et al. 2011).

El modelo obtenido logró la correcta clasificación de las muestras, tanto las adulteradas como las auténticas, con un 95% de confianza y un rendimiento analítico superior a las 120 muestras por hora (Nascimento, Araújo et al. 2011).

Los resultados derivados de los trabajos precedentes indican la posibilidad alcanzar los objetivos desde un inicio proyectados para esta investigación mediante la aplicación del análisis multivariante a los espectros UV-Vis, a pesar de ser una técnica sencilla.

Todos los laboratorios de criminalística del país cuentan con esta tecnología, por lo que no se requieren posteriores inversiones para generalizar los modelos y métodos que se deriven

de esta investigación. Es por ello que se decide emplear la espectrofotometría de absorción molecular UV-Vis como el primer paso hacia la clasificación automatizada de los rones.

Por lo tanto, a continuación se hará referencia a algunos aspectos teóricos de interés para la espectrofotometría UV-Vis y la quimiometría.

1.3. ESPECTROFOTOMETRÍA DE ABSORCIÓN MOLECULAR UV-VIS

La ley de Bouer – Lambert – Beer (*ecuación 1*), por usanza llamada ley de Beer, es el principio fundamental de la espectrofotometría de absorción molecular UV-Vis(Rubinson and Rubinson 2000; Rubinson and Rubinson 2001).

$$A = -\log T = \log {}^{P_0}/_{P} = \varepsilon bc \qquad (1)$$

Aunque pocas, se han encontrado excepciones a la generalización de que la absorbancia tiene relación lineal con el camino óptico. Estas irregularidades se originan por limitaciones propias de la ley de Beer, y por desviaciones que pueden ser de índole instrumental o de naturaleza química (Rubinson and Rubinson 2000; Rubinson and Rubinson 2001).

Las incertidumbres o ruidos asociados al instrumento suelen limitar la exactitud y la precisión de los análisis espectrofotométricos. A causa de que, en la mayoría de las medidas, el valor medio del ruido es constante e independiente de la magnitud de la señal es la relación señal/ruido un parámetro de calidad mucho más útil que el ruido para describir la calidad de un método analítico o el funcionamiento del espectrofotómetro. Esta relación se define como el inverso de la desviación estándar relativa, y como regla general se establece que debe ser mayor que dos o tres (Rubinson and Rubinson 2000; Rubinson and Rubinson 2001).

Para no incurrir en faltas, no se deben ignorar: el ruido instrumental como función de la transmitancia, otras fuentes de ruido instrumental como el producido por el disparo del detector de fotones y las incertidumbres en la posición de la cubeta, que pueden ser importantes en los espectrofotómetros UV-Vis de alta calidad. También se deben considerar el efecto de la anchura de la rendija en las medidas de absorbancia y el efecto de

radiación dispersada de longitudes de onda extremas de un instrumento (Pungor and Horvai 1994; Rubinson and Rubinson 2001; Ahuja and Jespersen 2006; Skoog, Holler et al. 2007).

Las medidas de absorción de la radiación UV-Vis tienen una gran aplicación en la identificación y determinación de una enorme cantidad de especies químicas orgánicas e inorgánicas. Sus aplicaciones en el análisis cualitativo están en cierta medida limitadas, ya que el número de máximos y mínimos de absorción es relativamente pequeño, por tanto, no siempre se puede realizar una identificación inequívoca (Rubinson and Rubinson 2001; Skoog, Holler et al. 2007).

1.4. QUIMIOMETRÍA

El acoplamiento de sistemas de registro de mayor eficacia a los instrumentos de medición y trabajo en los laboratorios permitió obtener una mayor cantidad de datos a un costo analítico relativamente bajo. Entonces surgió un nuevo dilema: cómo aprovechar al máximo e interpretar toda la información que estos nuevos sistemas ponen a disposición del analista. Cualquier cantidad de datos carece absolutamente de valor alguno hasta tanto no sea posible interpretarlos, la Quimiometría es la disciplina calificada para solventar estos problemas.

Massart uno de los fundadores de la Quimiometría la definió como *"la disciplina química que utiliza la Matemática, la Estadística y la Lógica formal para, diseñar o seleccionar procedimientos experimentales óptimos y proporcionar la máxima información química relevante a partir del análisis de datos químicos" (Massart 1999).*

Entre los artículos publicados en la revista *Analytical Chemistry* donde se aplican conocimientos de la Quimiometría se pueden identificar como principales áreas de trabajo el reconocimiento de patrones, la calibración multivariada, los estudios de QSAR, el procesamiento de señales químicas, la resolución matemática de mezclas complejas, el diseño de experimentos y el análisis de imágenes químicas (Lavine 1998; Lavine 2000; Ramis and Gracía 2001; Lavine and Workman 2002; Ferreira 2004).

Asumiendo los objetivos proyectados para este trabajo se aplica el reconocimiento de patrones.

1.4.1. RECONOCIMIENTO DE PATRONES

El reconocimiento de patrones figura entre las primeras líneas de trabajo que emprendió la Quimiometría (Singh 2005).

Identificar relaciones y vínculos entre objetos es un objetivo crucial del reconocimiento de patrones. En este caso particular, los objetos son muestras químicas o grupos de muestras previamente caracterizadas mediante análisis instrumental; bajo condiciones ideales que permitan su agrupamiento y clasificación según la similitud de sus espectros, cromatogramas u otra forma de registro de la data experimental(Singh 2005).

Sin distinción alguna, estas herramientas deben clasificar, de forma rápida y objetiva, nuevas muestras como miembros de un grupo determinado (Vandeginste, Health et al. 1985; Brereton 2002).

Para su desempeño, el reconocimiento de patrones comprende dos etapas: el análisis exploratorio de datos, y los métodos de clasificación supervisados y/o no supervisados (Singh 2005).

1.4.1.1. Análisis exploratorio de los datos

La exploración de datos se utiliza para dar notoriedad a la información contenida en una matriz de datos multidimensional. Tanto la observación directa de los datos como las herramientas simples de representación (histogramas, o gráficos de dispersión sobre pares de variables) son ineficaces para describir en toda su extensión las estructuras presentes.

Análisis de componentes principales

Este análisis es conocido por sus siglas en inglés como PCA, (*Principal Component Analysis*).

El PCA constituye una valiosa herramienta de exploración de los datos de esencia estadística y de proyección; maximiza la información de la varianza presente en los datos y la representa en el menor número de dimensiones posibles.

Las componentes principales: son ortogonales entre sí, se construyen como combinaciones lineales de las variables originales y se erigen como nuevas variables latentes que modelan las tendencias principales presentes en la nube de datos. Analizar la posición de las muestras en este nuevo sistema de coordenadas y el peso de las variables originales en los componentes principales permite descubrir la configuración de los diferentes grupos similares y las correlaciones entre ellos (Vandeginste, Health et al. 1985; Rock 1997; Ramis and Gracía 2001; Ebensen 2002).

Análisis jerárquico de clústeres

Se trata de una técnica con origen en la taxonomía numérica, se conoce por sus siglas en inglés como HCA (*Hierarchical Cluster Analysis*).

Los resultados son presentados en un diagrama con forma de árbol jerárquico conocido como dendrograma. Aquí, la compresión de dos ramas del árbol representa una gran similitud entre los objetos y se incluyen como opciones diferentes métodos de agrupamiento: el simple, el completo, el de la media, el del centroide, y el método Ward (Kowalski 1984; Vandeginste, Health et al. 1985; Ferreira 2004).

1.4.1.2. Métodos de clasificación

En un análisis clasificatorio se construyen modelos que según las características del objeto son capaces de pronosticar la pertenencia de dicho objeto a una categoría.

En la clasificación supervisada se cuenta con información *a priori* sobre la existencia de clases entre los objetos. Este tipo de clasificación se utilizó ampliamente en la Quimiometría a través de los años. Para construir un modelo de clasificación supervisada es necesario disponer de una muestra de objetos de clase y valores de las variables predictivas conocidas (Kowalski 1984).

La pertenencia de los objetos a las clases puede ser supuesta, o sea, puede tratarse de una hipótesis a comprobar. La asignación de los objetos a las clases debe ser exhaustiva: todos los objetos pertenecen a alguna clase; y mutuamente exclusiva: ningún objeto pertenece a más de una clase (Vandeginste, Health et al. 1985).

De manera general en una clasificación supervisada se siguen los siguientes pasos: modelación del set de entrenamiento, validación del modelo, mejoramiento de los datos si fuere necesario y finalmente la aplicación del modelo (Singh 2005).

Entre los métodos de clasificación sobresalen: *k* vecinos más cercanos (*k*-NN);el análisis discriminante sobre la regresión por mínimos cuadrados parciales (PLS-DA); modelación blanda por analogía de clases (SIMCA) y otro que en los últimos años ha tenido mucho auge, las máquinas de soporte vectorial (SVM) (Cristianini and Shawe-Taylor 2002; Singh 2005).

Método de los k-vecinos más cercanos

El método *k*-NN, es un tipo de modelado fuerte pues todas las muestras van a ser clasificadas en una clase determinada. Aunque el resultado no fuere uno de los más precisos *k*-NN intenta categorizar al desconocido basándose en su proximidad a las muestras ya categorizadas alrededor de él. En este algoritmo específicamente, la clase de un desconocido depende de la clase de sus *k* vecinos más cercanos y para ello se calcula la distancia: Euclidiana, Manhattan, Mahalanobis o la Minkowsky (Mack 1978; Ramis and Gracía 2001; Ferreira 2004).

Modelación blanda e independiente por analogía de clases

La SIMCA es una técnica de clasificación en la que se analiza la similitud entre los objetos en una clase o grupo en lugar de una conducta idéntica. Utiliza un análisis de componentes principales para desarrollar un modelo de cada grupo o clase dentro del conjunto de entrenamiento, esas clases se modelan por una o series de estructuras lineales, dependiendo del número de componentes requeridos para reproducir los datos de la clase (Rock 1997; Singh 2005).

Análisis discriminante sobre la regresión por mínimos cuadrados parciales

Esta técnica es una combinación de la regresión y la clasificación, cuyos resultados dan una medida cuantitativa de qué tanto pertenece una muestra a una determinada clase.

El algoritmo PLS-DA da la posibilidad de detectar directamente las variables responsables de cualquier tipo de clasificación, o sea, las direcciones de máxima variabilidad. Esto evita las variables no relacionadas con la respuesta sin necesidad de buscar direcciones discriminatorias por medio del análisis exploratorio subjetivo. En general, este es un algoritmo que, al igual que la SIMCA, parte de reducir la dimensión de las matrices y establecer las fronteras entre las clases. La clasificación de objetos en análisis discriminante se hace en función de sus probabilidades de pertenencia a una u otra clase. Este análisis es aplicable solo cuando se trabaja con cuatro o menos clases (Singh 2005).

Máquinas de soporte vectorial

Las SVM son una nueva clase de clasificadores que recién se incorporan en la labor de la Quimiometría. Utilizan una variedad de *kernels*: polinomiales, de base radial y Gaussianos. Son métodos muy robustos, poco dado al sobre-entrenamiento de los modelos y muy útiles en sistemas complejos no linealmente separables (Cristianini and Shawe-Taylor 2002). Son para la creación de funciones a partir de un set de datos con entrenamiento previo. Estas funciones pueden ser de clasificación y sus salidas son binarias, aunque también pueden generarse funciones de regresión (Statnikov, Aliferis et al. 2011).

1.4.2. SISTEMAS COMPUTACIONALES VINCULADOS AL RECONOCIMIENTO DE PATRONES

En el área de la Quimiometría existen tres *software* con un papel protagónico por las facilidades que brindan: *Pirouette*, *Unscrambler* y *PLS Toolbox* (Ebensen 2002; Singh 2005; Russell and Cohn 2012).

Pirouette es de fácil utilización y cuenta con una interfaz gráfica muy útil para la visualización de los resultados. Sin embargo, su precio es de 5600 USD, no oferta códigos

fuentes ni licencias para fines comerciales y no vende a Cuba, factores que constituyen sus principales desventajas (Pérez 2012).

Unscrambler combina de forma eficiente las técnicas de análisis estadístico y el mapeo multivariante para la interpretación de datos químicos. Sin embargo, la interfaz gráfica para la visualización de los resultados es menos asequible al usuario, vende a Cuba pero su precio es de 4500 euros y exige un pago de 500 euros anuales para la actualización de la licencia, no oferta códigos fuentes, ni licencias para fines comerciales (Pérez 2012).

PLS Toolbox es una colección de rutinas quimiométricas avanzadas que trabajan en el ambiente computacional de Matlab. Contiene herramientas que permiten la exploración de datos químicos, tareas de clasificación, análisis de regresión y la construcción de modelos de predicción. No obstante, se necesitan conocimientos avanzados de Matlab por parte del especialista químico. Se oferta licencia con fines docentes o académicos para uso de la institución que contrata pero, no incluye la licencia ni se distribuye con Matlab y no vende a Cuba (Pérez 2012).

Quimiometrix

Este es un producto cubano autónomo que constituye una herramienta potente y fácil de usar y permite la interacción entre los datos originales y todos los posibles modelos resultantes de su procesamiento (Ferreira 2004).

Quimiometrix *v* 2.0 incluye entre sus funcionalidades, un compendio de todas las rutinas quimiométricas contenidas en sus homólogos a nivel mundial, para la exploración, predicción, clasificación y calibración de los datos. Incluye además SVM, como una nueva variante para la clasificación y calibración, hecho que constituye una novedad en este tipo de sistema.

Todas estas opciones están contempladas en una interfaz gráfica que permite al usuario una interacción amigable con el *software* y una amplia gama de funcionalidades para la edición y el pre-procesamiento de los datos (Ferreira 2004).

CAPÍTULO 2

PARTE EXPERIMENTAL: ESPECTROFOTOMETRÍA UV-VIS Y QUIMIOMETRÍA

Tomando como premisa que al aplicar técnicas de reconocimiento de patrones a los espectros UV-Vis de los rones es posible autentificarlos o identificar tipo y marca; y condicionada por las ventajas que ofrecen estos métodos y técnicas se decidió utilizar en este estudio la espectrofotometría UV-Vis y la quimiometría para lograr los objetivos propuestos.

Entre las principales ventajas que justifican esta selección se destacan: el elevado rendimiento analítico dado por la baja complejidad y rapidez con que se puede procesar cada muestra; la reducción de los costes; el bajo consumo de reactivos químicos y de muestra y las grandes posibilidades de generalización a nivel nacional.

2.1. MATERIALES Y MÉTODOS

En correspondencia con las bondades de estos métodos, no se requiere de reactivos para la obtención de los espectros UV-Vis, solamente agua destilada.

2.1.1. TOMA DE MUESTRAS

En el muestreo se consideraron las producciones de las marcas cubanas Havana Club y Santero de las roneras San José[†] y Paraíso[‡] respectivamente.

Para escoger muestras representativas se necesitó muestrear varios lotes de cada variedad pues el procesamiento final del ron sigue una receta particular de cada fabricante que conlleva a una mezcla de rones añejos de diferentes años (Huertas, 2006). Estas

[†] Ronera ubicada en la ciudad San José, provincia Mayabeque.
[‡] Ronera ubicada en el poblado Tuinucú, provincia Sancti Spíritus.

mezclas son en gran parte las responsables de las características del sabor y aroma de cada ron y junto con varias etapas de la producción del ron en las que los procesos no ocurren estrictamente controlados pueden ser causa de incertidumbre asociada a la desviación de las características particulares dentro de una misma clase de ron (Avakasian 1990; González 2003; Valdés 2008).

En resumen, se muestrearon cinco lotes distintos de cada denominación, seleccionados aleatoriamente de las producciones comprendidas entre los años 2012 y 2014.

En total se dispuso de 80 muestras, 48 para el entrenamiento del modelo de clasificación y un grupo extra de 32 muestras para la validación.

Tabla 2-1. Descripción del muestreo.

Marca	Tipicidad	Cantidad de muestras para entrenamiento	Cantidad de muestras para validación	Total de muestras
Havana Club	Añejo Blanco	3	2	5
	Añejo 3 Años	3	2	5
	Añejo Especial	3	2	5
	Añejo Reserva	3	2	5
	Añejo 7 Años	3	2	5
	Añejo Ritual	3	2	5
	Selección de Maestros	3	2	5
	Añejo Máximo	3	2	5
Santero	Aguardiente Natural	3	2	5
	Silver Dry	3	2	5
	Añejo Blanco	3	2	5
	Añejo	3	2	5
	Palma Superior	3	2	5
	Añejo 5 Años	3	2	5
	Añejo Reserva	3	2	5
	Añejo 7 Años	3	2	5
Totales		**48**	**32**	**80**

2.1.2. MEDIOS: INSTRUMENTAL Y SOFTWARE

Se utilizaron medios generales del laboratorio y micropipetas ajustables de la firma Dragonmed de 100 a 1000 µL, hechas en China.

Para medir los espectros de absorbancia se utilizaron espectrofotómetros UV-Vis de doble haz, de la GBC Scientific Equipment, marca Cintra, modelo 101. Se utilizaron cubetas de cuarzo marca Hellma, con 10 mm de paso óptico. Este equipo tiene como registrador la versión 2.2 del *software* Cintral.

Para procesar los datos obtenidos de la espectrofotometría se utilizó el *software* Quimiometrix, versión 2.0, producido por el Centro de Aplicaciones de Tecnologías de Avanzada, CENATAV, radicado en La Habana, Cuba. Este *software* resultó particularmente útil en el tratamiento previo y el análisis exploratorio de los datos, así como en el desarrollo y la validación de los modelos de clasificación.

2.1.3. MÉTODOS

Para la modelación se utilizaron diferentes métodos: la espectrofotometría UV-Vis para la obtención de los datos, y la quimiometría y el reconocimiento de patrones para su procesamiento. Los pormenores se relacionan a continuación.

2.1.3.1. Registro de los espectros UV-Vis

Para seleccionar el método de preparación de las muestras de rones y el registro de los espectros UV-Vis se consideraron varias alternativas en métodos que previamente se han empleado en otras bebidas alcohólicas (Muñoz-Muñoz, Pichardo-Molina et al. 2010; Nascimento, Araújo et al. 2011; Martelo-Vidal, Domínguez-Agis et al. 2013).Se estudió entonces las condiciones bajo las cuales los espectros de todas las muestras aportaran información útil y no existiese interferencia producida por ruidos.

En general, el registro de los espectros UV-Vis se realizó en modo absorbancia, midiendo en la región comprendida entre 200 y 650 nm, a una velocidad de barrido de 400nm/s y un tamaño del paso de 0.160 nm. Se realizaron tres repeticiones de los espectros y se obtuvo el espectro medio de dichas repeticiones para minimizar el tamaño de la data. Se consideró

pertinente obtener la mayor cantidad de información posible para realizar luego el descarte de variables por su trascendencia en el poder de discriminación y el poder de modelación.

Para lograr condiciones óptimas en el registro de los espectros se seleccionaron los rones añejos 3 Años y 7 Años de la marca Havana Club, y el Aguardiente Natural de Caña de la marca Santero con el objetivo de analizar la calidad de los espectros de estos a diferentes proporciones volumen muestra - volumen de agua destilada.

Luego de obtener los espectros se exportan del formato **.scan* del *software* Cintral al formato **.csv* para así obtener los datos de puntos de la absorbancia en función de la longitud de onda. Este formato **.csv* es ejecutable con cualquier versión de Microsoft Excel.

2.1.3.2. Estudio cualitativo por HPLC

De las muestras se inyectaron 15 µL en un equipo HPLC SHIMADZU modelo LC-10A conformado por un detector DAD-Agilent 1100 (detección a 280 nm), bomba LC-10A horno CTO-10A, procesador de datos SCL-10AVP, autoinyector PerkinElmer Serie 200, desgasificador DGU 14-A,columna Lichrosorb–100 RP-18 (150 mm x 4.6 mm d.i., 5 µm), y como disolvente de elución se utilizó una mezcla de 8% acetonitrilo, 90% agua destilada y 2% ácido acético (González 2011).

El flujo fue de 1 mL/min y la temperatura de columna a 35ºC. Los resultados fueron comparados con los tiempos de retención registrados en la literatura bajo estas condiciones (González 2011).

2.1.3.3. Estudio de la repetibilidad instrumental

Para un análisis de repetibilidad instrumental, se obtienen tres espectros repetidosde una muestra representativa por cada denominación o tipo de ron. Estos espectros se registran manteniendo constantes las variables: condiciones experimentales, analista, laboratorio e instrumentos de medición.

A cada muestra seleccionada para este estudio de repetibilidad se le determina la media, la desviación estándar y el coeficiente de variación de los valores de absorbancia medidos

para cada longitud de onda. La media de estos valores nunca debe exceder el 5% de variabilidad (Hoel 1979; Brereton 2002; Miller and Miller 2002; Brereton 2003; Pérez 2012).

La media aritmética se determina por la ecuación:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2}$$

La desviación estándar se determina por la expresión:

$$S_x = \sqrt{\frac{\sum (x_i - x)^2}{n-1}} \quad ; \{i \in \mathbb{N};\ 1 \leq i \leq n\} \tag{3}$$

El coeficiente de variación se determina por la ecuación:

$$CV\% = \frac{S_x}{\bar{x}} \cdot 100 \tag{4}$$

2.1.3.4. Creación de las matrices de entrenamiento y validación

De los espectros ya exportados al formato **.csv*, se extraen los datos y se copian a una única hoja de cálculo de Excel donde se obtuvo una matriz donde cada término se relaciona con la absorbancia de cada muestra a una longitud de onda determinada, la cual posteriormente se importa al *software* Quimiometrix.

Para clasificarlos diferentes rones, es preciso contar con dos conjuntos de muestras, uno para realizar el entrenamiento y el otro, elaborado con muestras externas al modelo, para realizar la validación. Partiendo de este fundamento se confeccionan las matrices necesarias.

A partir de la base de datos preparada con los espectros UV, se escogen dos conjuntos de muestras y se elaboraron las matrices de entrenamiento y de validación, *E* y *V* respectivamente, donde *E* tiene dimensión [48; 2719] y *V*[32; 2719].

2.1.3.5. Aplicación de técnicas quimiométricas

Sobre la plataforma del *software* Quimiometrix v. 2.0 se realiza un análisis exploratorio de los datos con base en el PCA y el HCA. Luego para lograr el modelo de clasificación se utilizan los métodos *k*-NN, SIMCA y SVM.

Más adelante se hace un enfoque detallado de estos temas.

Análisis exploratorio de los datos

Para realizar el análisis exploratorio de los datos espectrales se realizó un PCA con el objetivo de representarlos datos en un espacio de dimensión menor al generado por los datos y al que estos se ajusten lo mejor posible; y transformar las variables originales, en general correlacionadas, en nuevas variables incorrelacionadas y ordenadas en orden de importancia de acuerdo a la variabilidad, o sea, determinar las llamadas *componentes principales*. En conjunto con el PCA se realizó además el descarte de las variables redundantes.

También se realizó un HCA para establecer una definición formal del esquema de clasificación.

CRITERIOS PARA LA SELECCIÓN DE VARIABLES:

Las variables que se utilicen en el modelo deben ser resultado del acuerdo entre el poder total de modelación y el poder de discriminación. Estos objetos dan una idea de la importancia de cada variable para la modelación de las clases.

En muchos problemas el poder de discriminación de la variable es muy importante para la interpretación de la estructura de los datos; en particular cuando han sido incluidas un número grande de variables sin información previa de su relevancia para el problema.

Cuando se reduce el número de variables no se debe hacer solo sobre la base del poder de discriminación de las mismas ya que esto puede convertirse en una sobreestimación de las diferencias entre las clases y el diseño de un método que busca las diferencias no debe estar condicionado a maximizar estas diferencias. La selección de variables resulta más objetiva

si se basa en ambos objetos, o sea, incluye también el poder de modelación, pues una variable con poco poder de modelación debe ser eliminada completamente.

Luego, se obtiene un resultado razonable eliminando solo las variables con bajo poder de modelación y discriminación.

El poder de modelación para una variable toma valores entre cero y uno (*ecuación 5*) (Massart 1999).

$$MP_j = 1 - \frac{S_{j,residual}}{S_{j,X}} \tag{5}$$

Siendo $S_{j,\ residual}$ la desviación estándar residual para la variable j en una clase q (*ecuación 6*), mientras que $S_{j,\ X}$ es la desviación estándar de la variable j (*ecuación 7*). Esta relación entre las desviaciones estándares es equivalente a la relación señal/ruido. Por lo tanto a medida que esta relación aumenta el contenido de información de la variable decrece.

$$S_{j,residual} = \sqrt{\frac{\sum_{p=0}^{I_q}\left(res_{pj}^{q}\right)^2}{I_q - A_q - 1}\cdot\frac{J}{J - A_q} + \frac{\sum_{p=1}^{I_r}\left(res_{pj}^{r}\right)^2}{I_r - A_r - 1}\cdot\frac{J}{J - A_r}} \tag{6}$$

$$S_{j,X} = \sqrt{\frac{\sum_{p=1}^{I_q}\left(x_{pj}^{q} - \bar{x}_j\right)^2}{I_q - 1} + \frac{\sum_{p=1}^{I_r}\left(x_{pj}^{r} - \bar{x}_j\right)^2}{I_r - 1}} \tag{7}$$

El poder de discriminación da una medida de la capacidad que tiene cada variable para lograr la separación entre las diferentes clases (*ecuación 8*) (Massart 1999).

$$\Phi_j^{r,q} = \sqrt{\frac{\left(S_{j,r}^{q}\right)^2 - \left(S_{j,q}^{r}\right)^2}{\left(S_{j,r}^{r}\right)^2 - \left(S_{j,q}^{q}\right)^2}} - 1 \tag{8}$$

Donde: $S_{j,r}^{q} = \sqrt{\frac{J\sum_{p=1}^{I_r}\left(res_{p,j}^{q}\right)^2}{I_r\left(J - A_q\right)}}$ (9)

Siendo $res^{q}_{p,j}$ para las muestras *p* de la clase *r* proyectadas en la clase *q*.

Un valor próximo a cero indica un bajo poder de discriminación, mientras que un valor superior a uno, cuanto mayor, mejor será el poder de discriminación.

El PCA se desarrolla en varias etapas considerando diferentes criterios:

EXAMEN DE LOS DATOS: A las datas importadas se les halla la matriz traspuesta para hacerla entendible al Quimiometrix, y se grafican los espectros UV-Vis de las muestras del conjunto de entrenamiento para detectar presencia de datos anómalos groseros o variables dudosas así como la existencia de grupos de muestras definidos.

TRANSFORMACIONES DE LOS DATOS: Para eliminar ruidos espectrales se realiza el alisamiento por la media móvil, con valor 10 en el parámetro ancho de ventana.

PREPROCESAMIENTO DE LOS DATOS: Se centran los datos con respecto a la media para evitar corrimientos en las proyecciones de los datos, logrando mantener los datos en un rango donde se toma como centro a la media de cada variable.

DIAGNÓSTICO O ANÁLISIS DE MUESTRAS ANÓMALAS: Al diagnóstico de datos anómalos se llega por tres medidas diferentes, indicadoras de cuánto una muestra se aproxima por *k* componentes principales. Tales medias son: los residuos de muestras, la distancia Mahalanobis y la probabilidad.

Aquellas muestras que caen fuera de uno de los dos umbrales son potencialmente anómalas. Estos datos anómalos, si existen, deben ser eliminados y se repite el proceso desde el inicio.

SELECCIÓN DEL NÚMERO DE FACTORES, OPCIONES DE VALIDACIÓN, MODO DE OPERACIÓN Y NIVEL DE CONFIANZA: Se trabajó considerando diez factores, y realizando validación por pasos, y exclusión de una muestra. Los cálculos se realizaron en modo de operación ligero y exhaustivo y un 95% de nivel de confianza como parámetro del test de Fisher.

SELECCIÓN DEL NÚMERO DE COMPONENTES PRINCIPALES: Se debe emplear el número correcto de factores o componentes principales que de acuerdo al porcentaje de varianza de cada componente principal, el porciento de varianza acumulada y el RMSECV brinde una descripción general de la data muy cercana a los valores reales, implique un menor costo de cómputo y permitan separar mejor las muestras de acuerdo a su clase.

Se visualizan (gráfico de *scores*) las muestras en las nuevas variables o componentes principales para analizar la separación en clases y los pesos de las variables (gráfico de *loadings*) para lograr interpretar los resultados. Además, se considera el poder de modelado para una posterior reducción de variables.

En el HCA, orientado a las muestras, se establecieron como parámetros la distancia Euclidiana y la unión por el método de Ward. Se realizó alisamiento por la media móvil a un ancho de ventana de 10 en caso de no haberlo realizado con anterioridad.

La elección del método de Ward está justificada por tener elevada eficiencia, tendencia a crear clústeres de pequeño tamaño, un carácter invariante bajo transformaciones monótonas de la matriz de distancias y sensibilidad a los *outliers* (Kowalski 1984; Rock 1997; Ebensen 2002).

Para la determinación del número de grupos existen diversos métodos: algunos están basados en intentar reconstruir la matriz de distancias original, otros en los coeficientes de concordancia de Kendall y otros realizan análisis de la varianza entre los grupos obtenidos. En tal sentido no existe un criterio universalmente aceptado, por lo tanto se comparará el agrupamiento obtenido considerando varias de estas normas para que la formación de los clústeres esté en acuerdo entre estas (Vandeginste, Health et al. 1985; Ferreira 2004).

Como uno de los criterios a considerar, se estableció un umbral de aproximadamente un 95% de similitud para los clústeres, no separándose en otras clases aquellos que fuesen similares en más del 95% (Singh 2005).

No obstante, dado que la mayor parte de los paquetes estadísticos proporciona las distancias de agrupamiento, es decir, las distancias a las que se forma cada grupo, se escogió también un método consistente en localizar en qué iteraciones del método utilizado dichas distancias los puntos se desvían de una línea recta. Se cuentan entonces el punto anterior al salto y los puntos siguientes.

Métodos de Clasificación

Una vez establecidas las clases, cada muestra se codifica y a cada clase se le asigna un número, con el cual se distinguirá la pertenencia de cada muestra a un tipo o marca de ron. Este código es ubicado en la última columna de la data en el Quimiometrix y se convierte al bloque C para llevar a cabo la clasificación.

Se pusieron a prueba los clasificadores siguientes:

***k*-NN:** Se aplica la distancia Euclidiana para los diez vecinos más cercanos. Se analiza la matriz para cada valor de *k*, el total de errores y los errores de clasificación, y se repite entonces la clasificación tomando en cuenta la mayor cantidad de vecinos que no conducen a errores (Mack 1978).

SIMCA: Los datos del conjunto de entrenamiento fueron clasificados empleando el número máximo de componentes principales, obtenidos en el PCA, sin validación y con validación cruzada, para un 95% de intervalo de confianza. Se analizó la proyección de las clases en los nuevos ejes establecidos por los componentes principales, los errores de clasificación y el poder de modelación y discriminación de las variables.

SVM: Los datos fueron clasificados empleando el *kernel* lineal y RBF. Se analizaron las clases asignadas a cada muestra y la probabilidad de pertenencia a la otra clase, seleccionando el que mejor resultado de complejidad brinde.

Selección del mejor modelo de clasificación

Habiéndose realizado la clasificación por las diferentes vías se determinan los errores de clasificación para cada clase y el error total sobre los conjuntos de entrenamiento y validación. Estos valores de error son de gran utilidad al seleccionar el modelo de

clasificación idóneo mediante el análisis de la eficiencia, para ello se determinó la sensibilidad y la selectividad.

La sensibilidad se expresó como la relación de la cantidad de muestras clasificadas correctamente en el modelo de entrenamiento entre el total de muestras empleadas para obtener el modelo; mientras que la selectividad se obtuvo dividiendo la cantidad de muestras clasificadas correctamente al validar el modelo entre la cantidad de muestras presentes en el conjunto de validación. Para una mejor interpretación de estos análisis, se asumieron sus correspondientes valores porcentuales.

CAPÍTULO 3

ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS EXPERIMENTALES

Siguiendo las técnicas y procedimientos experimentales descritos anteriormente se desarrolló el trabajo de laboratorio y los posteriores procesamiento e interpretación química de los resultados obtenidos.

3.1. REGISTRO DE LOS ESPECTROS UV-VIS

Con certeza, el ruido y la intensidad de la señal analítica son determinantes en la calidad del modelo y su poder de clasificación y diferenciación. Por lo tanto, se determinaron las condiciones de dilución bajo las cuales resultó mínimo el ruido y significativa la intensidad de la señal.

Para optimizar las condiciones se consideró una muestra de Havana Club Añejo 7 Años porque los valores de absorbancias máximas en su espectro UV-Vis resultaron aproximadamente intermedios entre los rones blancos, menos añejados, y los de alto grado de añejamiento como el Havana Club Añejo Máximo.

La relación volumen de muestra/volumen de agua destilada que mejor resultado mostró fue 180 μL: 3 mL. No obstante, los espectros de los rones claros obtenidos bajo estas condiciones no aportaron información útil por tener niveles de intensidad de absorbancia poco significativos y muy próximos a la línea base. Esto resulta consecuencia de estar dichas muestras diluidas en exceso.

Unir en un único modelo los rones claros y los oscuros haría más complejo el proceso de clasificación. Además, los rones claros solo se agruparían en una clase porque el ensanchamiento de las bandas de absorción producido por la dilución impediría diferenciar entre las marcas y tipicidades. Un modelo que solo informe si un ron es claro u oscuro es poco eficiente ya que este parámetro se puede distinguir a simple vista.

Se impuso la necesidad de búsqueda de nuevas condiciones para obtener los espectros de los rones claros y poco añejados, y se obtuvieron las condiciones óptimas para la relación 1 mL: 2 mL.

La principal consecuencia derivada del establecimiento de estas condiciones para registrar los espectros fue la necesidad de realizar dos modelos de clasificación: uno para los rones claros y otro para los oscuros. Esto se justifica porque al obtener solo un modelo con los espectros registrados bajo condiciones diferentes se pueden generar problemas de clasificación producidos por el efecto de la dilución.

Luego, al observar los espectros obtenidos bajo las condiciones ya establecidas se constató la presencia de ruidos espectrales. Sobre todo, resultó apreciable el efecto producido por la radiación dispersada a longitudes de ondas extremas del instrumento, que produjo la aparición de un falso máximo de absorción a valores de λ muy próximos a los 200 nm. Válido es señalar además que a partir de los 210 nm el etanol comenzó a tener absorción. Estas constituyeron razones suficientes para solo considerar en la elaboración de los modelos las λ superiores a 215 nm(Rubinson and Rubinson 2000; Skoog, Holler et al. 2007).

Atendiendo todas estas reformas propuestas se registraron los espectros (*figura 3-1*), se excluyeron las absorbancias medidas a menos de 215 nm por estar asociadas a ruidos químicos e instrumentales.

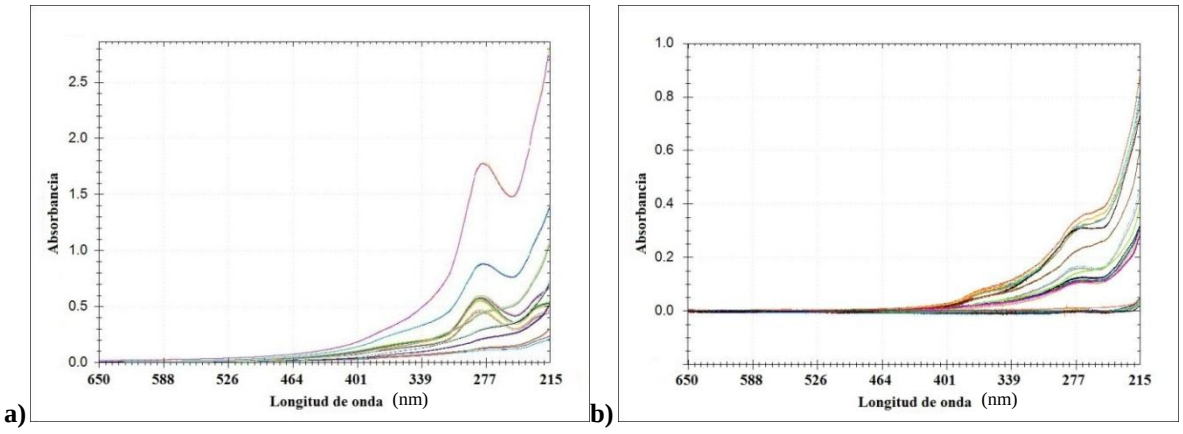

Figura 3-1. Espectros UV-Vis de las muestras de rones (a) oscuros y (b) claros.

3.1.1. ESTUDIO CUALITATIVO POR HPLC

Se obtuvieron los cromatogramas de dos tipicidades seleccionadas, el Aguardiente Natural de Santero y el Ritual Cubano de Havana Club. Al comparar estos cromatogramas con los tiempos de retención bajo las condiciones experimentales establecidas se puede observar la presencia de compuestos de la familia de los taninos, polifenoles y otros compuestos que se incorporan fundamentalmente en el añejamiento-acabado (*anexos 1 y 2*), sobre todo en el Ritual Cubano, el más añejo de los dos (González 2011).

El uso de DAD permitió comprobar que los picos correspondientes mostraron máximos de absorción alrededor de los 280 nm como característica significativa, hecho que indica su marcada influencia en las en el comportamiento de las intensidades de absorbancia de los rones analizados en la región UV. Este resultado se logró a pesar de obtenerse algunos picos con muy baja intensidad de la señal.

3.2. ESTUDIO DE LA REPETIBILIDAD INSTRUMENTAL

Se realizó el estudio correspondiente a la precisión del método mediante la repetibilidad instrumental. Se siguió el procedimiento descrito en el capítulo anterior (*epígrafe 2.1.3.3*) para los cálculos de las medidas de dispersión de los espectros representativos de cada tipicidad de los rones analizados. Los coeficientes de variación mostraron valores inferiores al 5% (*tabla 3-1*), lo que permitió comprobar la repetibilidad del método analítico bajo las condiciones de trabajo establecidas.

Tabla 3-1. Estudio de la repetibilidad instrumental para el registro de tres réplicas de los espectros de cada tipicidad de los rones objeto de estudio.

	Havana Club							
	Añejo Blanco	**Añejo 3 Años**	**Añejo Especial**	**Ritual Cubano**	**Añejo Reserva**	**Añejo 7 Años**	**S. de Maestros**	**Añejo Máximo**
Sx	0,00025	0,00027	0,00029	0,00024	0,00044	0,00034	0,00042	0,00194
CV%	4,25342	0,79364	0,16100	0,15917	0,33611	0,23377	0,11308	0,13048
	Santero							
	Aguardiente Natural	**Silver Dry**	**Añejo Blanco**	**Añejo**	**Palma Superior**	**Añejo 5 Años**	**Añejo Reserva**	**Añejo 7 Años**
Sx	0,00018	0,00017	0,00031	0,00025	0,00016	0,00022	0,00018	0,00022
CV%	1,83399	3,84405	1,23497	0,43377	0,24086	0,24795	0,15373	0,12898

3.3. ANÁLISIS EXPLORATORIO DE LOS DATOS

Se obtuvieron dos datas independientes. En efecto, se conformaron dos matrices de entrenamiento y dos matrices de validación que adquirieron dimensiones diferentes a las anteriores:

$$E_{oscuros}[33; 2719] \qquad E_{claros}[15; 2719]$$

$$V_{oscuros}[22; 2719] \qquad V_{claros}[10; 2719]$$

3.3.1. ANÁLISIS DE COMPONENTES PRINCIPALES PARA RONES OSCUROS

Se realizó un estudio preliminar de los datos observándose (*figura 3-1 a*) marcadas diferencias entre los espectros de las muestras de rones oscuros analizadas. Resultó apreciablemente mayor la intensidad en las absorbancias de los rones más añejos; y especialmente notable el máximo de absorción próximo a los 280 nm como rasgo particular de los rones Havana Club, principal elemento diferenciante con los de la marca Santero.

No se distinguieron datos anómalos o variables dudosas, mas sí se observó la existencia de grupos de muestras definidos.

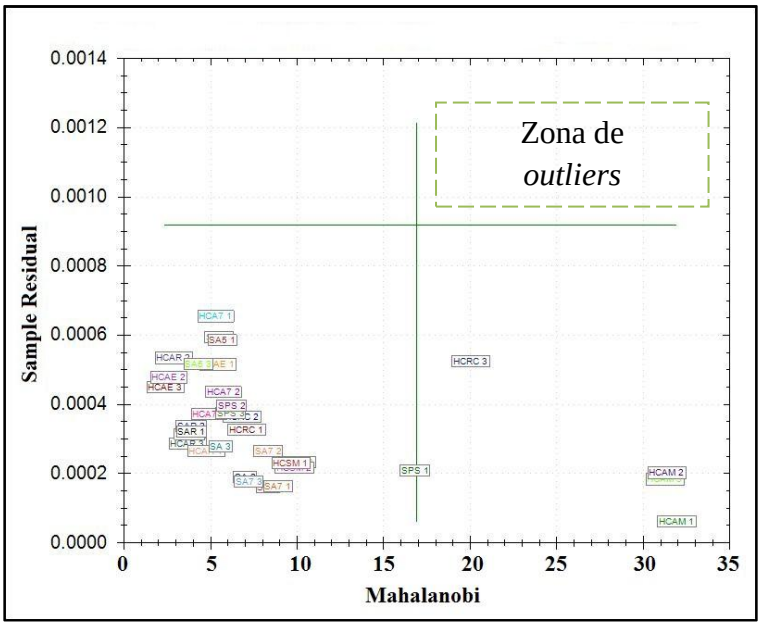

Figura 3-2. Análisis de *outliers* o datos anómalos en la modelación de rones oscuros.

En el test de *outliers* ninguna muestra se ubicó en el primer cuadrante, o sea no mostró resultados anómalos, por lo que no se hizo necesario desechar muestras ni repetir el PCA.

Tal como se observa en el gráfico de *outliers* (*figura 3-2*), son las muestras de Havana Club Añejo Máximo las que más se separaron de la nube de puntos, esto es en gran medida porque sus espectros tienen una intensidad de absorbancia mucho mayor que el resto de los rones y no precisamente como consecuencia de anomalías.

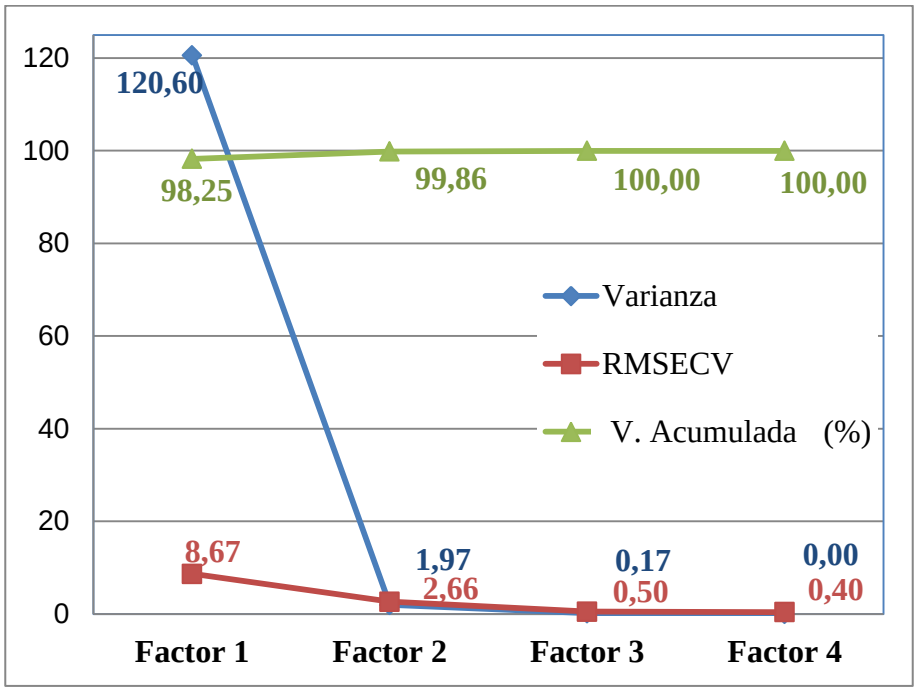

Figura 3-3. Análisis de la varianza en la modelación de rones oscuros.

Para la selección del número de componentes principales se hizo un estudio del comportamiento de la varianza, el error cuadrático medio y la varianza acumulada. Se analizó la varianza para el PCA realizado a la data de rones oscuros (*figura 3-3*), y se verificó que para un sistema que examine solo dos factores su valor es lo suficientemente bajo, 1.97. La varianza acumulada mostró que con dos componentes se explica el 99.86% de la variabilidad en la nube de datos y el valor RMSECV resultó muy bajo, 2.66.

En efecto, se demostró que con solo dos factores o componentes principales es posible obtener un modelo confiable y con un ínfimo costo de cómputo.

Tomando como criterio el peso de cada variable en los poderes de modelación y discriminación (*figura 3-4*), se excluyeron aquellas que aportaron información redundante y en consecuencia aumentaron el costo del cómputo de los datos y afectaron la calidad de la clasificación. En resultado, solo se incluyeron las variables comprendidas entre los 215 y 450 nm (*figura 3-5*) por lo que se robustecieron considerablemente los modelos.

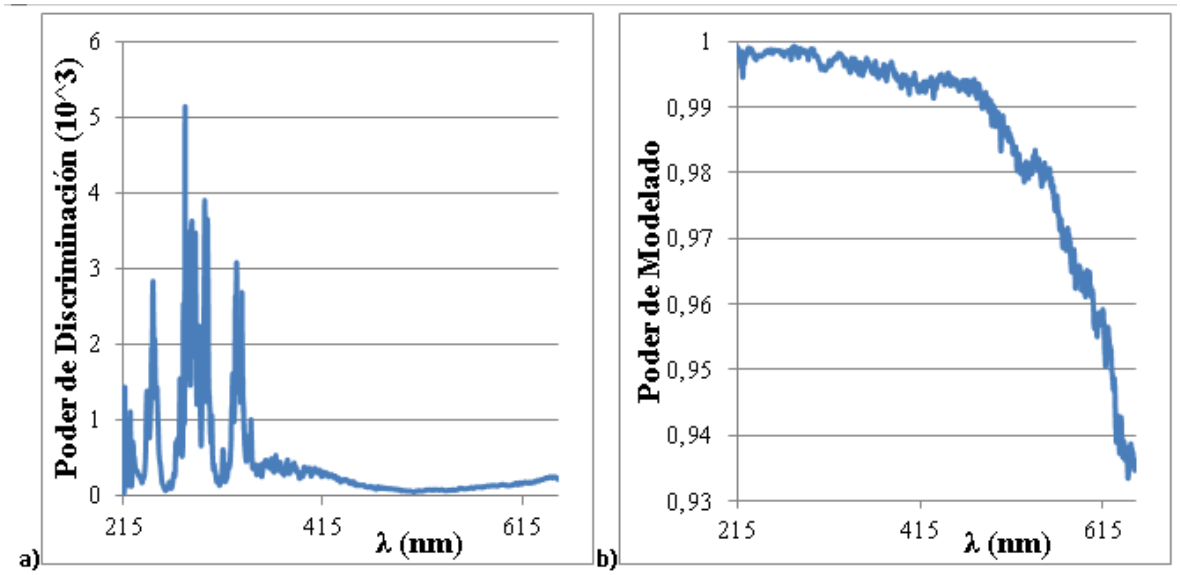

Figura 3-4. Selección de variables para el modelo de rones oscuros: (a) poder de discriminación y (b) poder de modelado.

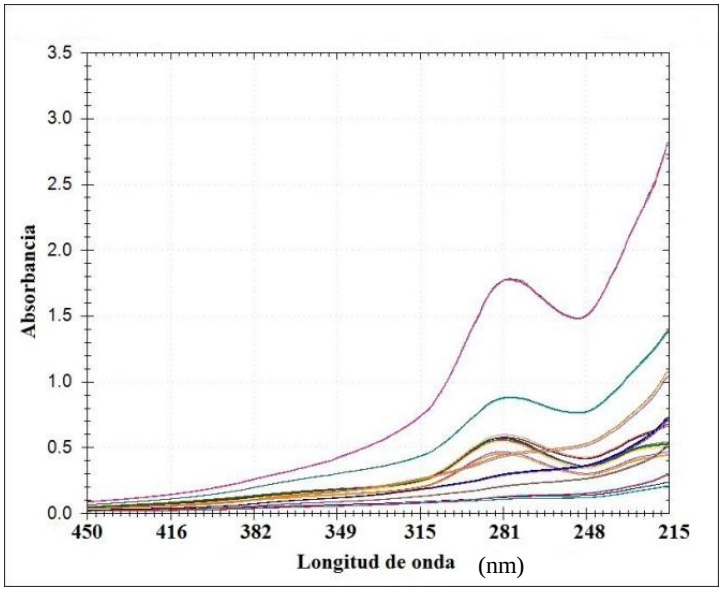

Figura 3-5. Espectros de rones oscuros luego de la selección de variables.

Estas variables se eliminaron de la data y en consecuencia se redimensionaron las matrices de entrenamiento y validación:

$$E_{oscuros}[33; 1469] \qquad V_{oscuros}[22; 1469]$$

El examen de las coordenadas de las muestras en los dos ejes de componentes principales seleccionados mostró tendencia a la formación de nueve grupos (*figura 3-6*), dos menos que los once posibles por la cantidad de variedades de ron procesadas. Esto reveló la posibilidad de que cuatro tipos de rones se encuentren agrupados en pares y no se puedan separar en clases distintas para la clasificación.

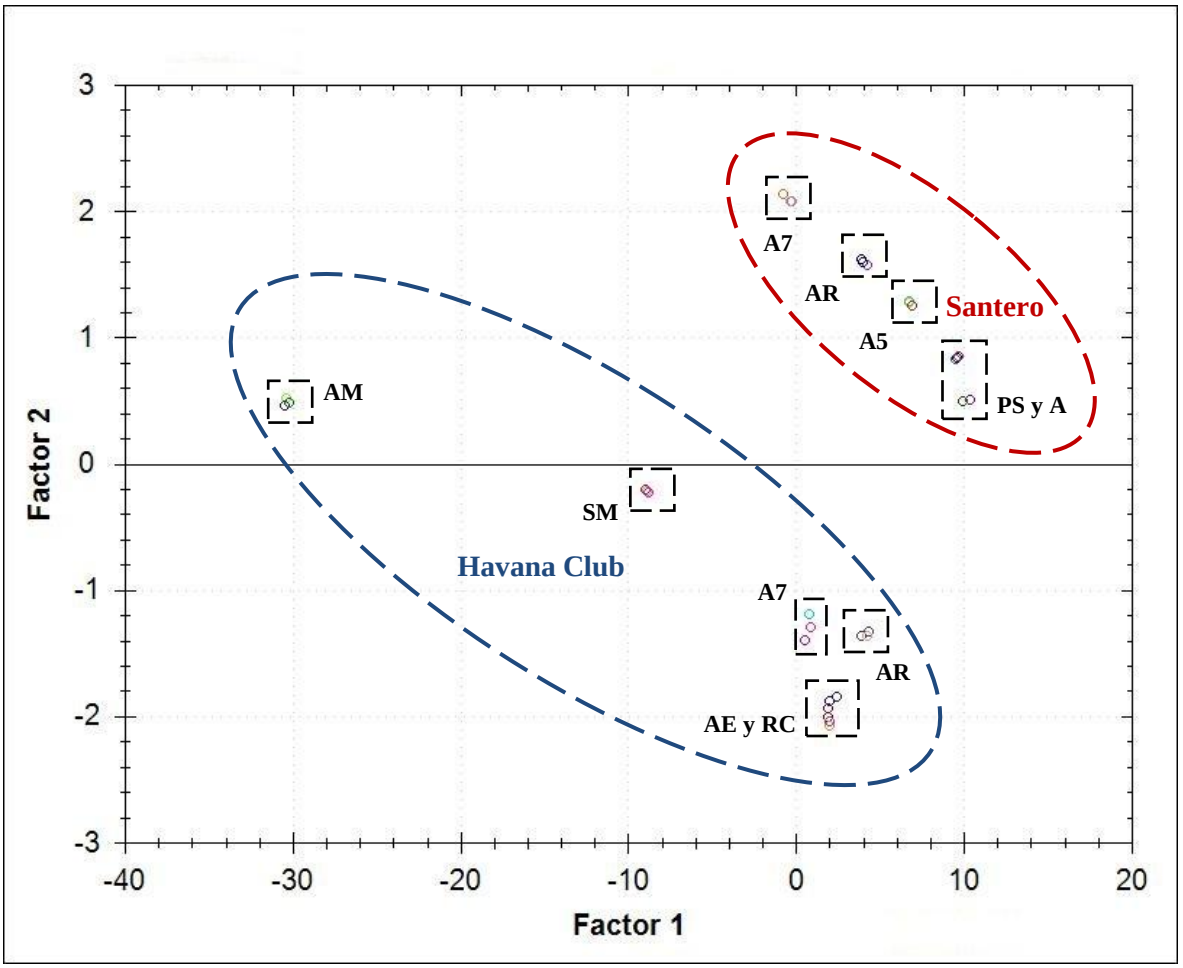

Figura 3-6. *Scores* para el modelo de rones oscuros.

Resultó de especial relevancia la clara tendencia al agrupamiento entre los rones de una misma marca ubicándose los Santero desde el primer cuadrante hasta el segundo en orden

de los menos a los más añejos; y los Havana Club desde el cuarto cuadrante, hasta el segundo pasando por el tercero, también en orden de los menos a los más añejos.

El primer componente principal, o factor 1, guarda relación con el añejamiento. Este hecho que depende de los niveles de intensidad de las absorbancias se evidencia por el orden de derecha a izquierda de los rones menos a los más añejos. Se explica así por qué el Añejo Máximo se ubica mucho más a la izquierda que el resto de los rones.

El segundo componente, está más vinculado a separación de las bandas de absorción y la presencia de compuestos de estructura química más compleja y caracterizada por la presencia de grupos auxócromos, anillos aromáticos y una marcada conjugación. Inmediatamente se impone pensar en los taninos y polifenoles que aporta la madera durante el añejamiento y en las especias u otros agregados colorantes o saborizantes que se incorporan a la formulación en la etapa de acabado y procesamiento final del ron. Este segundo factor es el que permite diferenciar mejor entre ambas marcas (Robinson and Company 1974; Gauglitz and Vo-Dinh 2006).

En paralelo, un análisis de la preponderancia de las variables predictivas originales en las combinaciones lineales formadas por las componentes principales mostró que longitudes de onda menores de 295 nm son las de mayor influencia en el factor 1. En consecuencia, los rones con mayores valores de absorbancia a estas longitudes de onda se ubican más a la izquierda en el gráfico de *scores*, hecho afín con el comportamiento de los rones más añejos y ratifica la idea de conectar esta primera componente con el añejamiento.

Con respecto al factor 2 se observaron dos tendencias principales, las longitudes de onda menores de 250 nm promueven un desplazamiento de los *scores* en sentido positivo, hacia arriba. Este efecto se agudiza por debajo de los 228 nm, de ahí que aquellos rones con altas intensidades de sus absorbancias a estas longitudes de onda tiendan a ubicarse en la región septentrional del gráfico. En contraposición, las λ entre 267 y 305 nm producen un efecto negativo en esta segunda componente, es decir, aquellos rones con mayor absorbancia alrededor de estas longitudes de onda muestran tendencia a localizarse en la región meridional del gráfico de *scores*. Estos hechos son consecuentes con el comportamiento típico de los rones Havana Club, que presentan un máximo de absorción aproximadamente

a los 280 nm y todos, a excepción del Añejo Máximo, cumplen con lo antes planteado. Por tanto, esta característica es la más preponderante en la diferenciación entre ambas marcas.

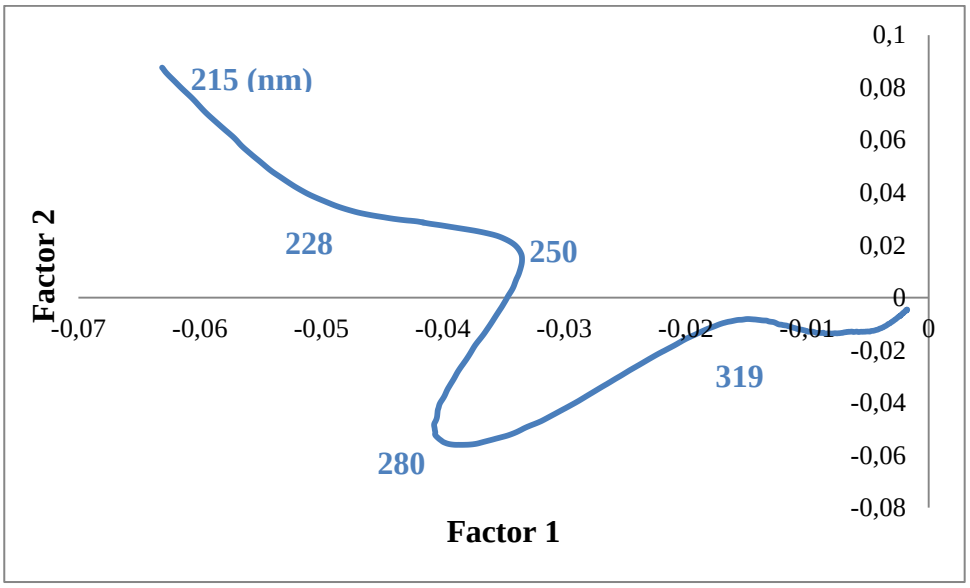

Figura 3-7. Gráfico de *loadings* o preponderancia de las variables predictivas originales en las combinaciones lineales formadas por las componentes principales del modelo de rones oscuros.

El comportamiento mostrado por el Havana Club Añejo Máximo no debe verse como un hecho insólito.

En el gráfico de los *loadings* (*figura 3-7*) es perfectamente visible que las variables contribuyentes a los corrimientos positivos son superiores en número y ponderación que las causantes de los corrimientos negativos.

3.3.2. ANÁLISIS JERÁRQUICO DE CLÚSTERES PARA RONES OSCUROS

Se realizó el HCA. El dendrograma, para un 95% de similitud, formó nueve clústeres donde se agruparon las tres muestras de cada tipo dentro del mismo clúster. Igual resultado se obtuvo analizando las variaciones de las distancias en cada etapa (*figura 3-8*).

Estos resultados reafirmaron los obtenidos del PCA, por lo que se decidió formar nueve clases para la clasificación de rones oscuros (*tabla 3-2*).

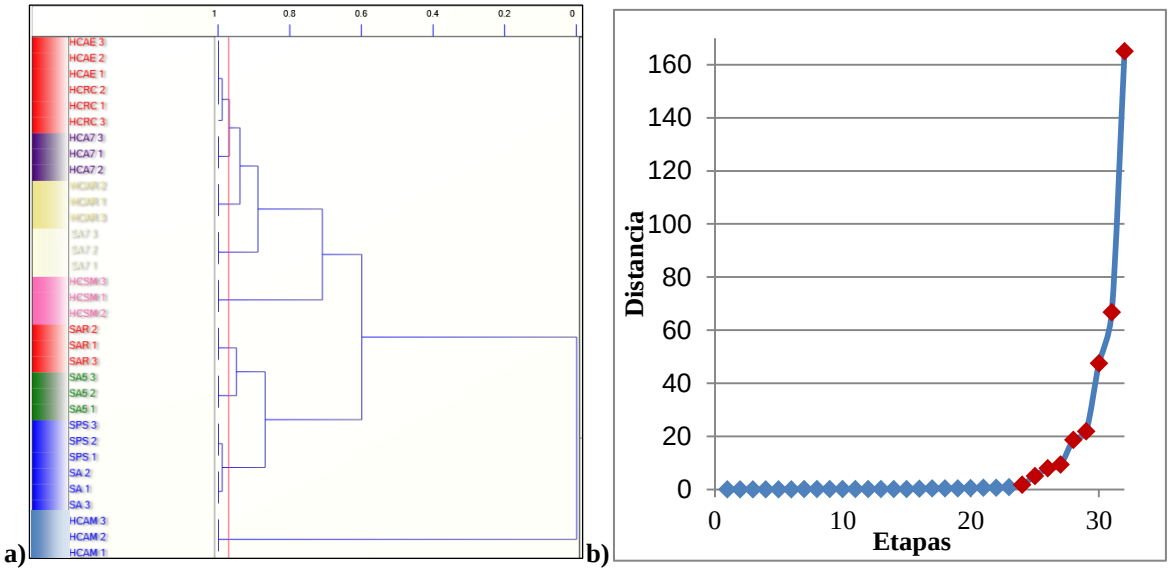

Figura 3-8. HCA para el modelo de rones oscuros: (a) dendrograma y (b) gráfico de distancia contra etapas.

3.3.3. ANÁLISIS DE COMPONENTES PRINCIPALES PARA RONES CLAROS

En un estudio inicial de la data de los rones claros sometidos al análisis se observaron (*figura 3-1 b*) diferencias notables en los espectros. El Aguardiente Natural de Caña y el Silver Dry, ambos de Santero, mostraron bajos niveles de intensidad de absorbancia;esto resulta consecuencia del empleo de caldos poco añejados para su fabricación, y por lo tanto, los taninos, polifenoles y otros componentes de la madera se van a hallar en bajas proporciones. Análogo a los rones oscuros, se apreció una mayor intensidad en las absorbancias como característica de los rones más añejos, pero no se observó el máximo a los 280 nm que distinguió los rones oscuros Havana Club de los Santero.

No se detectaron datos anómalos o variables dudosas pero sí se percibió variabilidad entre espectros de los Añejo Blanco de ambas marcas. Esto permitió inferir la posibilidad de dificultades en la modelación con respecto a esta tipicidad.

En el test de *outliers* (*figura 3-9*) todas las muestras se agruparon alejadas de la zona de datos anómalos, por lo que no se desechó ninguna y se continuó con el PCA.

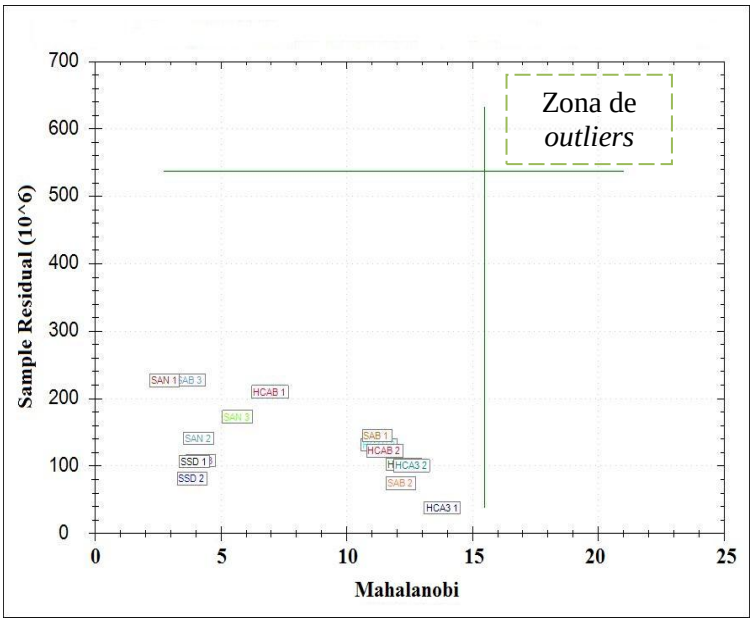

Figura 3-9. Análisis de *outliers* en la modelación de rones claros.

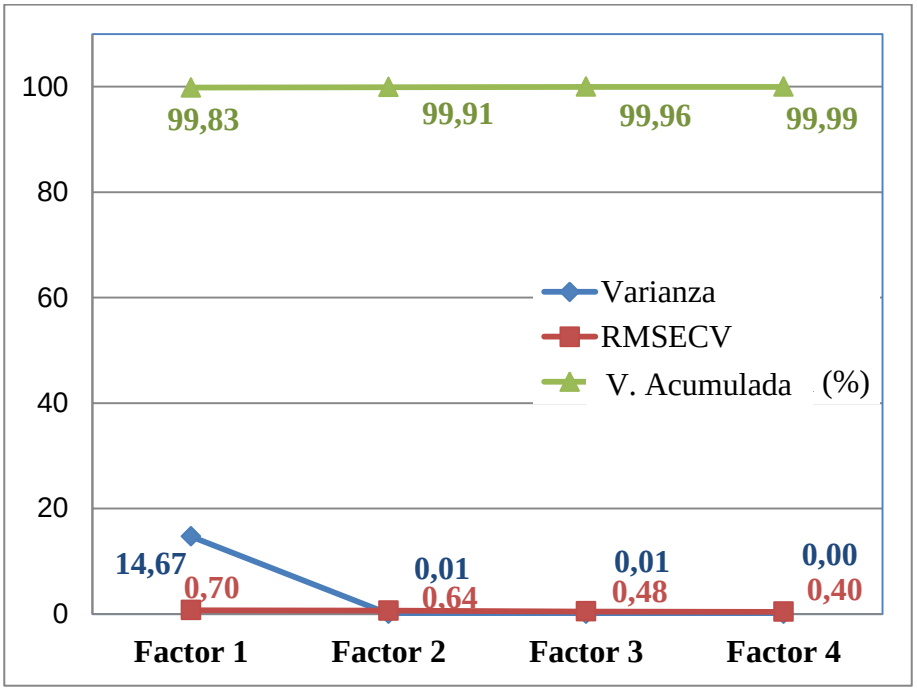

Figura 3-10. Análisis de la varianza en la modelación de rones claros.

Para la selección del número de componentes principales se hizo un estudio del comportamiento de la varianza, el error cuadrático medio y la varianza acumulada. En el análisis de la varianza para el PCA realizado a la data de rones claros (*figura 3-10*), se

constató que para un sistema que examine solo dos factores el valor de varianza en los datos es lo suficientemente bajo, 0.01. Además, la varianza acumulada expresó que con dos componentes se explica el 99.91% de la variabilidad en los datos y el valor RMSECV resultó muy bajo, 0.64. Este resultado demostró que con solo dos componentes principales se puede obtener un modelo viable que requiera un mínimo costo de cómputo en el procesamiento de los datos.

Al valorar el peso de cada variable en los poderes de modelación y discriminación, se despreciaron nuevamente las que aportaron información redundante, aumentaron el costo computacional y disminuyeron la calidad de la clasificación (*figura 3-11*).

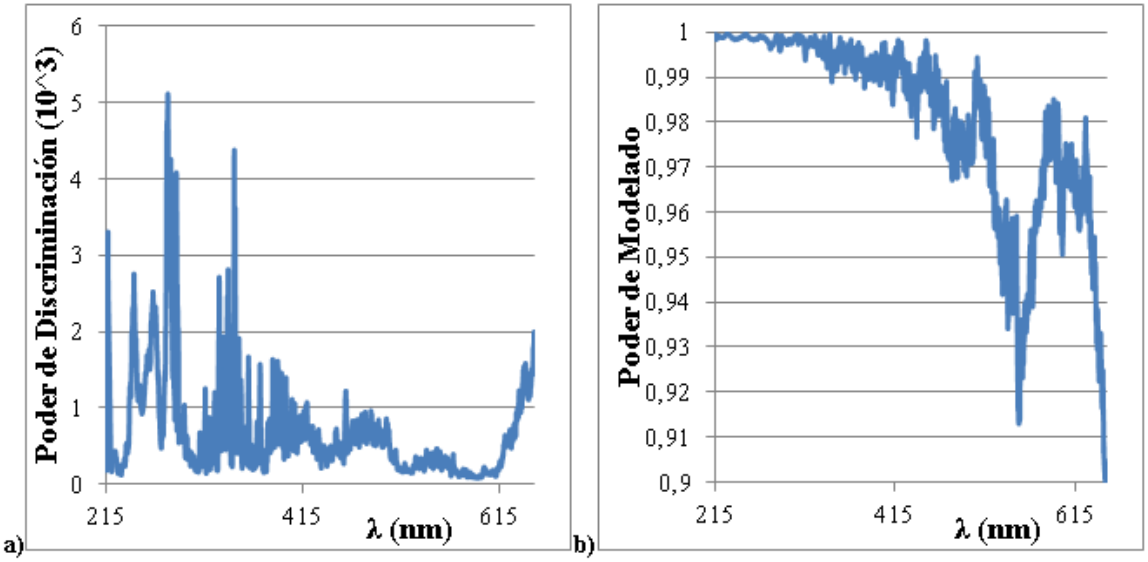

Figura 3-11. Selección de variables para el modelo de rones claros: (a) poder de discriminación y (b) poder de modelación.

Se observó (*figura 3-11*) que resultaba viable desechar las longitudes de onda mayores de 450 nm. Estas variables se eliminaron de la data y en consecuencia se redimensionaron las matrices de entrenamiento y validación:

$$E_{claros}[15; 1469]\ V_{claros}[10; 1469]$$

El descarte de variables redundantes favoreció en alto grado la clasificación, pues en el modelo de rones claros permitió diferenciar mejor los Añejo Blanco de ambas marcas.

Ambos modelos se robustecieron ganando en calidad y rapidez de procesamiento (*figura 3-12*).

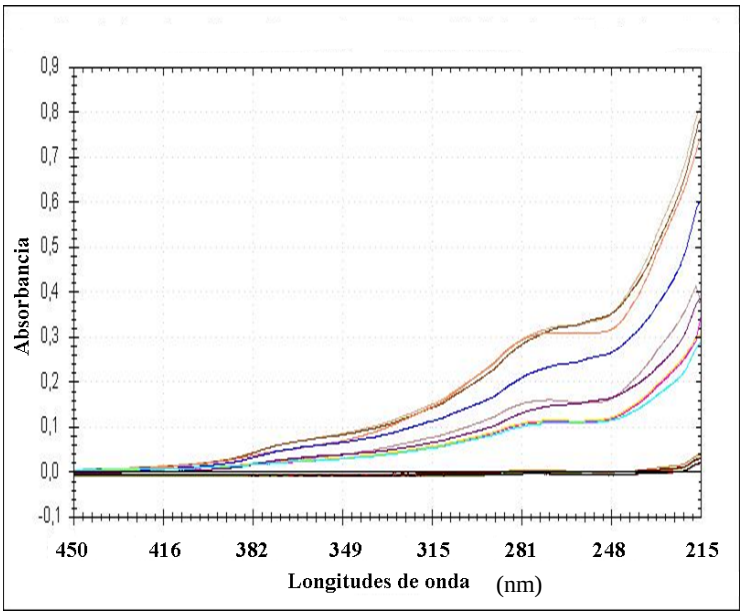

Figura 3-12. Espectros de rones claros luego de la selección de variables.

Al estudiar la distribución de los *scores* para las dos primeras componentes principales (*figura 3-13*) se observó disposición a formar cinco grupos, máxima cantidad esperada de acuerdo con el número de variedades procesadas. Esta buena separación constituye una mejora propiciada por la selección de variables, pues al considerar las 2719 variables iniciales se incluía información redundante de poca utilidad en la modelación y en la discriminación.

Similar a lo apreciado para los oscuros, se observó tendencia al agrupamiento entre los rones de una misma marca ubicándose los Santero desde el cuarto cuadrante hasta el segundo pasando por el primero en orden de los menos a los más añejos; y los Havana Club se distribuyeron angularmente desde las proximidades del origen de coordenadas hacia los cuadrantes segundo y tercero, también en orden de los menos a los más añejos.

En comparación con los *scores* obtenidos para los rones oscuros se observó mayor dispersión en las nubes de puntos de un mismo grupo, inducida esta por la marcada

variabilidad en las características propias de cada lote. Esta inconstancia se atribuye al uso de caldos de diferente añejamiento para las mezclas.

Resultó notable la relación entre el factor 1 y el grado de añejamiento, similar a los rones oscuros. Se evidenció una ordenación de derecha a izquierda de los rones menos a los más añejos, igualmente propiciada por los niveles de intensidad de las absorbancias de las muestras.

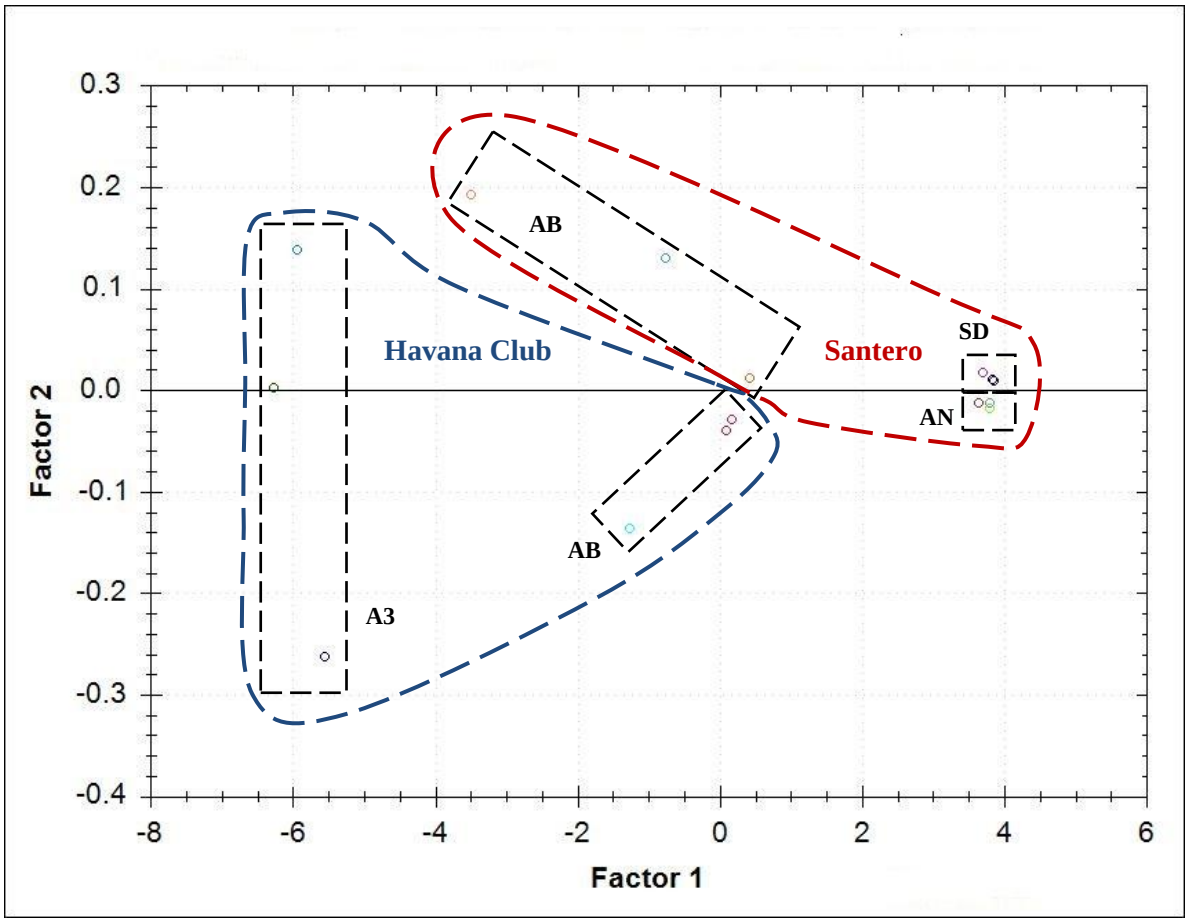

Figura 3-13. *Scores* para el modelo de rones claros.

Fue curioso cómo los Añejo Blanco se ubicaron, respecto a la primera componente, en una misma región, aún a pesar de tener el Santero una ligera tonalidad dorada y el Havana Club ser incoloro. Esto indica que la coloración del Santero se debe más a un aditivo colorante para hacerlo comercialmente más llamativo; indica también que el color no es tan preponderante como los componentes incluidos en el formulado durante el añejamiento.

La segunda componente mantiene vínculo con la presencia de compuestos aromáticos, con marcada conjugación en las insaturaciones y grupos auxócromos que producen corrimientos batocrómicos y efectos hipercrómicos en los máximos de absorción. Estos efectos están justificados por los mismos motivos que los rones oscuros (*epígrafe 3.3.2*) y, nuevamente, marcan la diferencia entre el Havana Club y el Santero. No obstante esta diferenciación es más incipiente porque los compuestos activos a estas longitudes de onda están más diluidos y la sensibilidad de las bandas de absorción disminuye por efecto de la polaridad del disolvente.

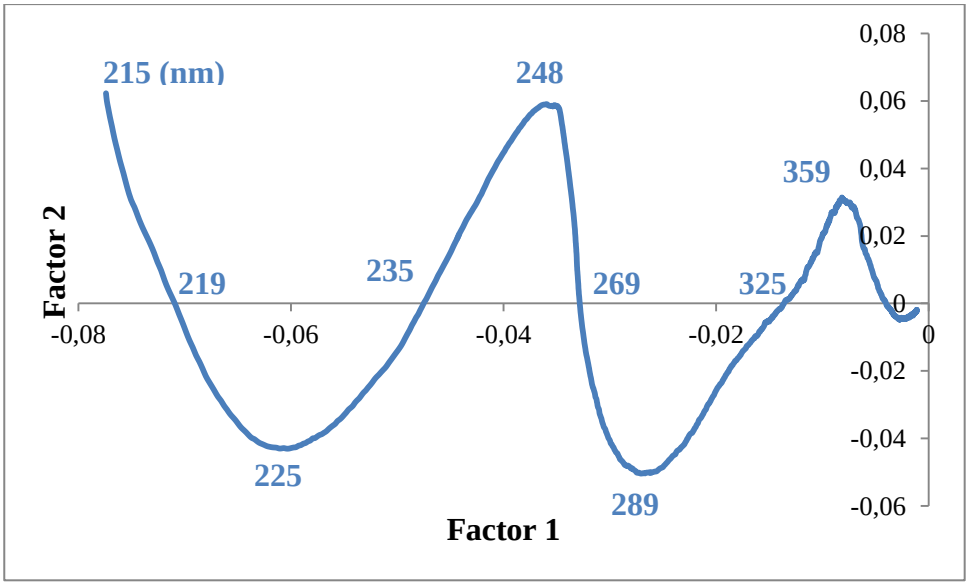

Figura 3-14. Gráfico de *loadings* o preponderancia de las variables predictivas originales en las combinaciones lineales formadas por las componentes principales del modelo de rones claros.

Luego, en el examen de la ponderación de las variables predictivas originales en las combinaciones lineales formadas por los factores 1 y 2 se comprobó que las longitudes de onda menores de 289 nm ejercen mayor influencia sobre el factor 1. Consecuentemente, los rones con intensidad de la absorbancia a estas longitudes de onda se ubican más a la izquierda en los *scores*, similar a lo ocurrido con los rones oscuros (*figura 3-14*).

Con respecto al factor 2, se constató que las λ en las que los espectros experimentaban una inflexión cóncava propician una orientación meridional en el gráfico de *scores* y que las

inflexiones convexas propician la orientación septentrional. En resumen, esta componente está más ligada a la relación entre mínimos y máximos en la absorción por lo que se deduce que con disolventes menos polares es probable obtener una mejor modelación, ya que se gana sensibilidad ante los máximos y mínimos de absorción al no influir la polaridad del agua(Robinson and Company 1974; Gauglitz and Vo-Dinh 2006).

Nuevamente fue el segundo factor el que permitió distinguir las marcas, aun siendo casi imperceptible la presencia de mínimos y máximos de absorción por el efecto asociado a la dilución y la polaridad del disolvente.

3.3.4. ANÁLISIS JERÁRQUICO DE CLÚSTERES PARA RONES CLAROS

El dendrograma obtenido como resultado del HCA, para un 95% de similitud, formó cinco clústeres, pero a consecuencia de la variabilidad entre los lotes no se agruparon las tres muestras de cada tipo dentro del mismo clúster. La muestra 2 de Santero Añejo Blanco apareció en un clúster independiente y más cercano al Havana Club de 3 años. Tampoco se agruparon correctamente el resto de los Añejo Blanco, y los aguardientes y Silver Dry se unieron en un mismo clúster (*figura 3-15*).

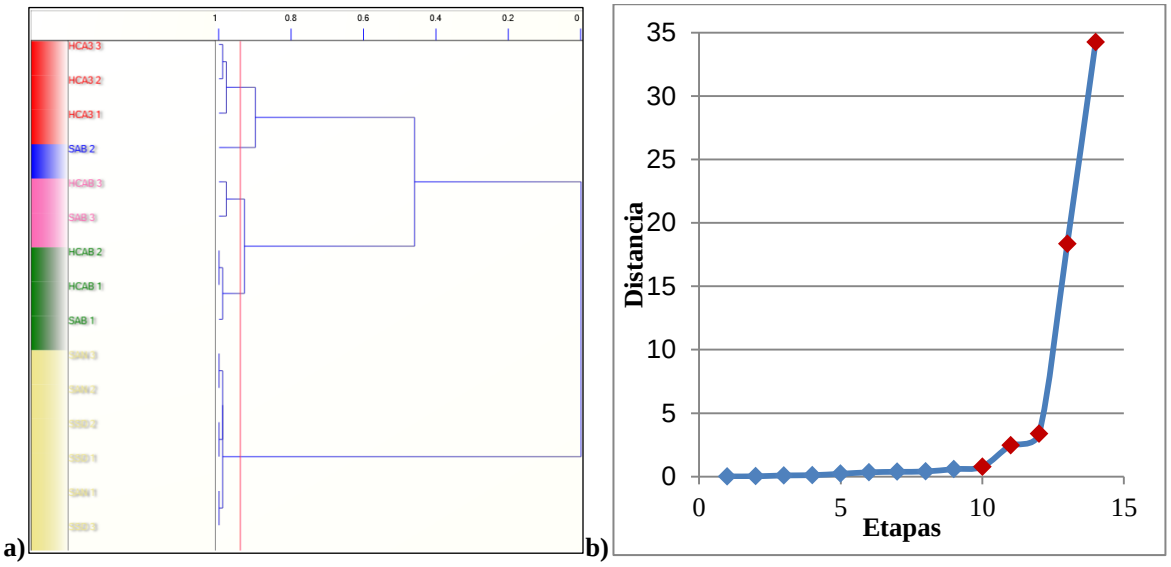

Figura 3-15. HCA para el modelo de rones oscuros: (a) dendrograma y (b) gráfico de distancia contra etapas.

Estos resultados discreparon con los derivados del PCA, que ofrece una visión más realista porque pondera equilibradamente los dos factores, y por lo tanto se hizo prevalecer su criterio. Analizando las variaciones de las distancias en cada etapa se dedujo también la formación de cinco grupos de conglomerados.

No obstante, este desacuerdo entre el PCA y el HCA es indicio de cuán compleja puede resultar la clasificación de los rones claros. Esta idea está reforzada también por la elevada variabilidad entre diferentes lotes de una misma tipicidad.

3.3.5. SELECCIÓN Y FORMACIÓN DE LAS CLASES

Como resultado del análisis exploratorio realizado a la data, con los rones oscuros se decidió formar nueve clases de once posibles como máximo, mientras que para los rones claros se formaron las cinco posibles, ya que existió buena separación entre los clústeres (*tablas 3-2 y 3-3*).

Tabla 3-2. Descripción de las clases en el modelo de rones oscuros.

Clase No.	Tipos de rones que contiene cada clase	Cantidad de muestras por clase
1	Santero Añejo Reserva	2
2	Santero Añejo 5 Años	2
3	Santero Añejo 7 Años	2
4	Santero Añejo y Santero Palma Superior	4
5	Havana Club Añejo Especial y Havana Club Ritual	4
6	Havana Club Añejo Reserva	2
6	Havana Club Añejo 7 Años	2
8	Havana Club Selección de Maestros	2
9	Havana Club Añejo Máximo	2

Tabla 3-3. Descripción de las clases en el modelo de rones claros.

Clase No.	Tipos de rones que contiene cada clase	Cantidad de muestras por clase
1	Santero Silver Dry	2
2	Santero Aguardiente Natural de Caña	2
3	Santero Añejo Blanco	2
4	Havana Club Añejo Blanco	2
5	Havana Club Añejo 3 Años	2

3.4. ENTRENAMIENTO Y VALIDACIÓN DE LOS MODELOS DE CLASIFICACIÓN

Para entrenar los modelos de clasificación se aplicaron los algoritmos *k*-NN, SIMCA, y SVM. Durante la modelación se aplicó la validación cruzada con exclusión de una muestra para garantizar la calidad de los mismos.

3.4.1. MODELACIÓN PARA RONES OSCUROS

Al aplicar a la matriz de entrenamiento el algoritmo *k*-NN mediante la distancia Euclidiana y 4 vecinos más cercanos no hubo errores en la clasificación (*tabla 3-4*).

Tabla 3-4. Matriz de confusión para el entrenamiento de rones oscuros por *k*-NN.

Clases	1	2	3	4	5	6	7	8	9
1	3	0	0	0	0	0	0	0	0
2	0	3	0	0	0	0	0	0	0
3	0	0	3	0	0	0	0	0	0
4	0	0	0	6	0	0	0	0	0
5	0	0	0	0	6	0	0	0	0
6	0	0	0	0	0	3	0	0	0
7	0	0	0	0	0	0	3	0	0
8	0	0	0	0	0	0	0	3	0
9	0	0	0	0	0	0	0	0	3

La clasificación por el SIMCA tampoco mostró errores. Además, se obtuvo buena sensibilidad, precisión y especificidad en el análisis.

Tabla 3-5. Matriz de confusión para el entrenamiento de rones oscuros por SIMCA.

Clases	1	2	3	4	5	6	7	8	9	No clasifica
1	3	0	0	0	0	0	0	0	0	0
2	0	3	0	0	0	0	0	0	0	0
3	0	0	3	0	0	0	0	0	0	0
4	0	0	0	6	0	0	0	0	0	0
5	0	0	0	0	6	0	0	0	0	0
6	0	0	0	0	0	3	0	0	0	0
7	0	0	0	0	0	0	3	0	0	0
8	0	0	0	0	0	0	0	3	0	0
9	0	0	0	0	0	0	0	0	3	0

A través del algoritmo de SVM empleando un kernel lineal y aplicando *Grid Search*[§], se obtuvo 1 y 0 como valores ideales para los parámetros *C* y *Rho* respectivamente. Bajo estas condiciones la clasificación no tuvo errores (*tabla 3-6*).

Tabla 3-6. Matriz de confusión para el entrenamiento de rones oscuros por SVM.

Clases	1	2	3	4	5	6	7	8	9
1	3	0	0	0	0	0	0	0	0
2	0	3	0	0	0	0	0	0	0
3	0	0	3	0	0	0	0	0	0
4	0	0	0	6	0	0	0	0	0
5	0	0	0	0	6	0	0	0	0
6	0	0	0	0	0	3	0	0	0
7	0	0	0	0	0	0	3	0	0
8	0	0	0	0	0	0	0	3	0
9	0	0	0	0	0	0	0	0	3

3.4.2. VALIDACIÓN Y SELECCIÓN DEL MODELO DE RONES OSCUROS

Se validaron los modelos obtenidos por los algoritmos *k*-NN, SIMCA y SVM. En resultado, los modelos *k*-NN y SVM no realizaron clasificaciones equívocas, mientras que el modelo SIMCA tuvo 20 aciertos de las 22 muestras a clasificar.

Para seleccionar el mejor modelo de clasificación se tomó en cuenta la sensibilidad y selectividad de cada uno de los modelos.

Los resultados impusieron una disyuntiva, elegir el modelo *k*-NN o el SVM ya que ambos mostraron sensibilidad y selectividad máxima (*tabla 3-7*) El clasificador SVM es más potente y robusto que el *k*-NN y además, reduce la dimensión de la data al basarse en vectores soporte para efectuar la clasificación. Estas razones justificaron la elección del modelo SVM para efectuar la clasificación de los rones oscuros (Mack 1978; Cristianini and Shawe-Taylor 2002; Statnikov, Aliferis et al. 2011).

[§] Herramienta que permite la selección de los parámetros del modelo de forma automática a través de un proceso de optimización por validación cruzada. Este paso solía hacerse por el método de prueba y error.

Tabla 3-7. Análisis de sensibilidad y selectividad de los modelos de rones oscuros.

Modelo	Errores de Clasificación	Errores de Validación	Sensibilidad (%)	Selectividad (%)
***k*-NN**	0	0	100	100
SIMCA	0	2	100	90.91
SVM	0	0	100	100

3.4.3. MODELACIÓN PARA RONES CLAROS

Tabla 3-8. Matriz de confusión para el entrenamiento de rones claros por *k*-NN.

Clases	1	2	3	4	5
1	2	1	0	0	0
2	1	2	0	0	0
3	0	0	0	2	1
4	0	0	1	2	0
5	0	0	0	0	3

Al aplicar a la matriz de entrenamiento el algoritmo *k*-NN mediante la distancia Euclidiana y 2 vecinos más cercanos este cometió varios errores en la clasificación (*tabla 3-8*). Los errores se deben a que el algoritmo *k*-NN, por su mayor aporte a la modelación, hizo prevalecer el criterio de la primera componente sobre el de la segunda (Singh 2005). Una mejor óptica de este planteamiento se tiene si se analiza en conjunto la tabla de errores y el gráfico de *scores* (*tabla 3-8 y figura 3-13*)

La clasificación por el SIMCA no mostró errores. Además, se obtuvo buena sensibilidad, precisión y especificidad en el análisis (*tabla 3-9*).

Tabla 3-9. Matriz de confusión para el entrenamiento de rones claros por SIMCA.

Clases	1	2	3	4	5	No clasifica
1	2	1	0	0	0	0
2	0	2	0	0	1	0
3	0	0	1	0	2	0
4	0	0	0	2	1	0
5	0	0	0	0	3	0

A través del algoritmo de SVM, y el clasificador C_SVC empleando un *kernel* lineal y aplicando *Grid Search*, se obtuvo 112 y 0 como valores ideales para los parámetros *C* y *Rho* respectivamente. No se encontraron errores de clasificación bajo este régimen de términos (*tabla 3-10*).

Tabla 3-10. Matriz de confusión para el entrenamiento de rones claros por SVM.

Clases	1	2	3	4	5
1	3	0	0	0	0
2	0	3	0	0	0
3	0	0	3	0	0
4	0	0	0	3	0
5	0	0	0	0	3

3.4.4. VALIDACIÓN DEL MODELO DE RONES CLAROS

Se realizó la validación de los modelos obtenidos por los algoritmos *k*-NN, SIMCA y SVM. En todos los modelos se obtuvieron clasificaciones erradas: En resultado, los modelos obtenidos por *k*-NN y SVM no realizaron clasificaciones equívocas, mientras que el modelo SIMCA tuvo solo 7 aciertos de las 22 muestras a clasificar.

Para seleccionar el mejor modelo de clasificación se tomó en cuenta la sensibilidad y selectividad de cada uno de los modelos (*tabla 3-11*).

El algoritmo SVM, por tener potencia y robustez superior al resto, demostró ser el más sensible y selectivo de los modelos entrenados (Cristianini and Shawe-Taylor 2002). Por ende, el algoritmo SVM es el más idóneo para lograr la clasificación de los rones claros. El SIMCA, resultó poco sensible y selectivo, razón suficiente para no ser tomado en cuenta.

Tabla 3-11. Análisis de sensibilidad y selectividad de los modelos de rones claros.

Modelo	Errores de Clasificación	Errores de Validación	Sensibilidad (%)	Selectividad (%)
***k*-NN**	6	2	60	80
SIMCA	5	7	66.67	30
SVM	0	1	100	90

Los resultados obtenidos en este modelo demostraron que para lograr una clasificación acertada para los rones claros es necesario realizar un estudio más exhaustivo. Se debe incrementar en número de muestras para el entrenamiento y la validación porque entre los diferentes lotes existe una elevada variabilidad de las propiedades de absorción UV-Vis. Incluso, debe estudiarse la posibilidad de emplear otras técnicas y métodos analíticos porque la dilución a la que se encuentran las sustancias activas en el UV produce ensanchamiento de las bandas de absorción y disminuye el poder de discriminación en los modelos.

CONCLUSIONES

1. Se desarrollaron los modelos de clasificación automática de los rones objeto de estudio por marcas y tipos a partir de sus espectros UV-Vis y la aplicación de técnicas quimiométricas de reconocimiento de patrones.
2. Se estableció la medición de absorbancia en el intervalo de 215 a 450 nm, y a diluciones de 180 µL en 3 mL de agua destilada para el registro de los rones oscuros y 1 mL en 2 mL para los claros.
3. Las determinaciones cualitativas por HPLC-DAD indicaron que los taninos y polifenoles determinan el comportamiento de la absorción de los rones en la región UV y hacen posible la diferenciación entre las dos marcas y los diferentes tipos de ron.
4. Los estudios de selección de variables permitieron reducir las dimensiones de las matrices de entrenamiento y validación, se facilitó así el cómputo de los datos y se eliminó la información redundante que afectaba la modelación.
5. La exploración de los datos indicó la formación de nueve clases para la modelación de los rones oscuros, existiendo un adecuado agrupamiento y orientación de los clústeres mientras quelos rones claros, aunque existió tendencia al agrupamiento, no se evidenció buena orientación en los clústerescomo consecuencia del bajo grado de añejamiento y las variaciones de producción entre lotes.
6. Los mejores resultados en el desarrollo y validación de los modelos de clasificación se obtuvieron para los entrenados con las SVM, que mostraron un 100% de sensibilidad tanto en rones claros como en oscuros y una selectividad de 90 y 100% respectivamente.

RECOMENDACIONES

1. Incrementar el número de muestras de lotes representativos por cada tipicidad tanto en los sets de entrenamiento como en los de validación, lo que permitirá robustecer los modelos.
2. Valorar el empleo de otras técnicas y métodos analíticos para la modelación de rones claros.
3. Analizar la influencia de los diferentes modos de adulteraciones en los espectros UV-Vis de las muestras.
4. Estudiar la factibilidad de la preconcentración y extracción en fase apolar de las muestras para obtener mayor información de los espectros UV-Vis.

BIBLIOGRAFÍA

Ahuja, S. and N. Jespersen (2006). Modern Instrumental Analysis, Elsevier Science.

Avakasian, S. P. (1990). Teoría general de añejamiento. Conferencia Internacional de Ciencia y Tecnología de los Alimentos, La Habana.

Banguela, Y. (2001). Determinación del Tiempo de Permanganato (Prueba de Lang) en destilados de rones. Licenciatura en Química. Santa Clara, Universidad Central Marta Abreu de Las Villas.

Bevin, C. J., A. J. Fergusson, et al. (2006). "Development of a Rapid "Fingerprinting" System for Wine Authenticity by Mid-infrared Spectroscopy." Journal of Agricultural and Food Chemistry **54**(26): 9713-9718.

Brereton, G. (2002). Data Analysis for the Laboratory and Chemical Plant. Nueva York, John Wiley & Sons Inc.

Brereton, R. G. (2003). Chemometrics: Data Analysis for the Laboratory and Chemical Plant, Wiley.

Cardoso, D. R., L. G. Andrade-Sobrinho, et al. (2004). "Comparison between Cachaça and Rum Using Pattern Recognition Methods." Journal of Agricultural and Food Chemistry **52**(11): 3429-3433.

Cenzano, J. M. (2003). Análisis de Vinos, Mostos Y Alcoholes, Ediciones Mundi-Prensa.

Comité Técnico de Normalización, N. C., de Bebidas alcohólicas, (2001). Ron. Especificaciones. NC 113. O. N. d. Normalización. La Habana, Oficina Nacional de Normalización. **113**.

Comité Técnico de Normalización, N. C., de Bebidas alcohólicas, (2007). Bebidas alcohólicas. Determinación de componentes volátiles mayoritarios en bebidas alcohólicas destiladas, aguardientes y alcohol etílico por cromatografía gas-líquido NC 508. C. T. d. N. N. C. d. B. alcohólicas. La Habana, Comité Técnico de Normalización.

Cordoví, E. (1988). Bebidas notables. Santiago de Cuba, Editorial Oriente.

Corominas, J. (1954). Ron. Diccionario crítico etimológico de la Lengua Castellana. Madrid, Editorial Gredos. **IV**.

Cristianini, N. and J. Shawe-Taylor (2002). "An Introduction to Support Vector Machines." Cambridge University Press.

Cubadebate (2012) Havanista, una respuesta cubana al caso Havana Club. Cubadebate

Díaz, F. V. (2010). Enología: vinos, aguardientes y licores, Publicaciones Vértice SL.

Dromi, J. R. and MERCOSUR (1996). Código del MERCOSUR: tratado, protocolos, acuerdos, declaraciones, decisiones, resoluciones, directivas, recomendaciones, Ciudad Argentina.

Ebensen, K. (2002). Multivariate Data Analysis in Practice. Oslo, Camo Process As. .

Ferreira, M. (2004). Quimiometría. Introducción a la Quimiometría. CENATAV, La Habana. **2a Edición**.

Gauglitz, G. and T. Vo-Dinh (2006). Handbook of Spectroscopy, Wiley.

González-Arjona, D., G. López-Pérez, et al. (2006). "Supervised Pattern Recognition Procedures for Discrimination of Whiskeys from Gas Chromatography/Mass Spectrometry Congener Analysis." Journal of Agricultural and Food Chemistry **54**(6): 1982-1989.

González, J. C. (2011). El roble usado en Cuba para el añejamiento del ron: características físico-químicas, sensoriales y anatómicas. Departamento de limentos. La Habana, Universidad de La Habana. **Licenciado**.

González, Y. (2003). Análisis del Añejamiento tradicional y acelerado de materias primas que se usan para la formulación del Ron Mulata. Ingeniería Química. Santa Clara, Universidad Central Marta Abreu de Las Villas. **Ingeniero**.

Hoel, P. G. (1979). Estadística Elemental. La Habana, Edición Revolucionaria.

Huertas, J. (2006). "FERMENTACIÓN Y BIOTECNOLOGÍA." TMAG 4992. Retrieved junio 27, 2012, from http://www.javierhuertas.com/PTMA-04-09.html.

Iglesias, F. R. (2003). Proyecto Andalucía: Antropología : economía y technología, Publicaciones Comunitarias.

Johll, M. E. (2008). Química e investigación criminal: Una perspectiva de la ciencia forense, Reverté.

Klages, F. (1968). Tratado de química orgánica, Reverté.

Kowalski, B. R. (1984). Chemometrics: Mathematics and Statistics in Chemistry, Springer.

Lachenmeier, D. W., E.-M. Sohnius, et al. (2006). "Quantification of Selected Volatile Constituents and Anions in Mexican Agave Spirits (Tequila, Mezcal, Sotol, Bacanora)." Journal of Agricultural and Food Chemistry **54**(11): 3911-3915.

Lavine, K. (1998). "Chemometrics." Analytical Chemistry **70**: 209-228.

Lavine, K. (2000). "Chemometrics." Analytical Chemistry **72**: 91R-97R.

Lavine, K. and J. Workman (2002). "Chemometrics." Analytical Chemistry **74**: 2763-2770.

Mack, Y. (1978). K-Nearest neighbor estimation, University of California.

Marcano, J. E. (2010). "El Ron: Su nombre e historia." Mi país. La República Dominicana. Retrieved junio 22, 2012.

Markechova, D., P. Majek, et al. (2014). "Determination of the adulterants in adulterant-brandy blends using fluorescence spectroscopy and multivariate methods." Analytical Methods **6**(2): 379-386.

Martelo-Vidal, M. J., F. Domínguez-Agis, et al. (2013). "Ultraviolet/visible/near-infrared spectral analysis and chemometric tools for the discrimination of wines between subzones inside a controlled designation of origin: a case study of Rías Baixas." Australian Journal of Grape and Wine Research **19**(1): 62-67.

Massart, B. e. a. (1999). Handbook of Chemometrics and Qualimetrics. Amsterdam, Elsevier Sience.

Miller, J. N. and J. C. Miller (2002). Estadística y Quimiometría en Química Analítica. Madrid, Prentice Hall.

Muñoz-Muñoz, A. C., J. L. Pichardo-Molina, et al. (2010). "Identification and quantification of furanic compounds in tequila and mezcal using spectroscopy and chemometric methods." Journal of the Brazilian Chemical Society **21**: 1077-1087.

Nascimento, E. C. L., M. C. U. Araújo, et al. (2011). "A flow-batch analyzer for UV-Vis spectrophotometric detection of adulteration in distilled spirits." Journal of the Brazilian Chemical Society **22**: 1061-1067.

Ochoa, R. d. l. C. (2003). Delitos conta la economía nacional. Derecho penal especial. La Habana, Editorial Félix Varela. **2**.

Pascual, J. M. P. (2001). HOSTELERÍA Técnicas y Calidad de Servicio, Ediciones Hotel.

Paturau, J. M. (1989). By-Products of the Cane Sugar Industry: An Introduction to Their Industrial Utilization, Elsevier.

Paz, L. A. S. (2005). Determinación de metanol en bebidas alcohólicas fermentadas tradicionales y populares de mayor consumo en dos regiones de la República de

Guatemala por cromatografía de gases. Facultad de Ciencias Químicas y Farmacia. Guatemala, San Carlos de Guatemala. **Licenciado en química Farmacéutica**.

Pérez, E. R. (2012). Desarrollo de modelos quimiométricos de clasificación de los combustibles líquidos destilados del petróleo mediante el empleo de las espectroscopias ultravioleta e infrarroja. La Habana, Universidad de La Habana. **Licenciado**.

Picque, D., P. Lieben, et al. (2006). "Discrimination of Cognacs and Other Distilled Drinks by Mid-infrared Spectropscopy." Journal of Agricultural and Food Chemistry **54**(15): 5220-5226.

Piola, J. C. (2011). "Dietilenglicol y metanol: unidos por el mecanismo de toxicidad y por los brotes masivos." Retrieved 24 de abril, 2012.

Pungor, E. and G. Horvai (1994). A Practical Guide to Instrumental Analysis, Taylor & Francis.

Ramis, G. and M. C. Gracía (2001). Quimiometría. España, Editorial Síntesis SA.

Rivero, D. and M. C. Bertot, Eds. (2009). Código penal de la República de Cuba, Ley No. 62/87 (Anotado con las Disposiciones del CGTSP). La Habana, Ediciones ONBC.

Robinson, J. W. and C. R. Company (1974). Handbook of spectroscopy, CRC Press.

Rock, B. (1997). An Introduction to Chemometrics. General Areas.

Rodríguez, M. H. and A. S. Gallego (1999). Tratado de Nutrición, Díaz de Santos.

Rubinson, K. A. and J. F. Rubinson (2000). Contemporary instrumental analysis, Prentice Hall.

Rubinson, K. A. and J. F. Rubinson (2001). Análisis instrumental. Madrid, Pearson Educación. SA.

Russell, J. and R. Cohn (2012). The Unscrambler, Book on Demand.

Simunovic, Y. (1999). Manual de bebidas alcohólicas y vinagres. Colección de Manuales Jurídicos. S. A. y. Ganadero. Chile, Ministerio de Agricultura.

Singh, S. (2005). Pattern Recognition and Data Mining: Third International Conference on Advances in Pattern Recognition, ICAR 2005, Bath, UK, August 22-25, 2005, Springer.

Skoog, D. A., F. J. Holler, et al. (2007). Principles of Instrumental Analysis, Thomson Brooks/Cole.

Statnikov, A., C. F. Aliferis, et al. (2011). A Gentle Introduction to Support Vector Machines in Biomedicine: Theory and methods, World Scientific.

Sun, D.-W. (2008). Modern Techniques for Food Authentication. Canada, Academic Press, Elsevier.

Tamayo, J. O. (2009) En aumento las patentes de Cuba en Estados Unidos. El Nuevo Herald

Torre, N. R. d. l. (2010). "El Ron cubano: Origen, tipo y proceso de elaboración." Retrieved junio 27, 2012, from http://www.alambiques.com/ron_cubano.htm.

Tortosa, G. (2011). Manual práctico de Quimiometría, Digital.CSIC.

Valdés, T. (2008). Estimación de la Incertidumbre debida al efecto de matriz mediante clasificación de materiales. Licenciatura en Química. Santa Clara, Universidad Central Marta Abreu de Las Villas. **Licenciado:** 117.

Vandeginste, B., Health, et al. (1985). Chemometry: The Brains Behind Analysis.

A.1. CROMATOGRAMAS OBTENIDOS POR HPLC PARA EL HAVANA CLUB RITUAL CUBANO Y EL SANTERO AGUARDIENTE NATURAL

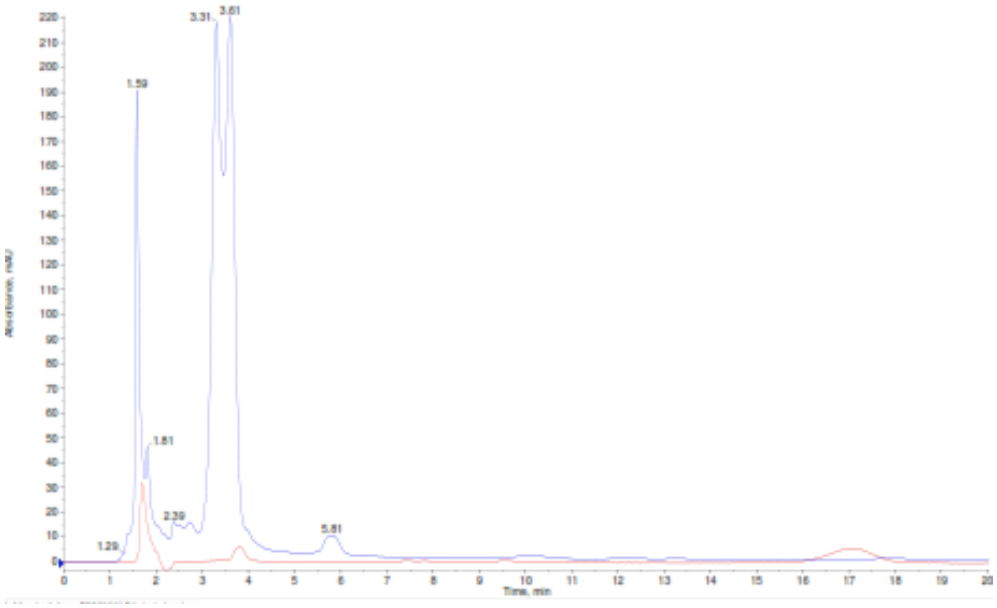

Figura 0-1. Cromatogramas superpuestos de ambos rones.

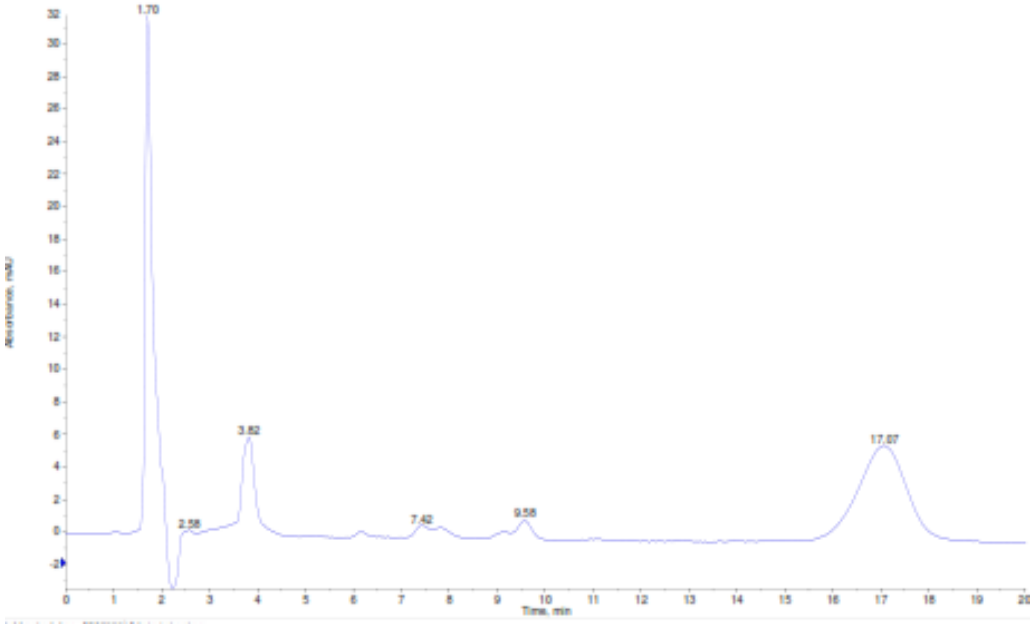

Figura 0-2. Cromatograma del Santero Aguardiente Natural.

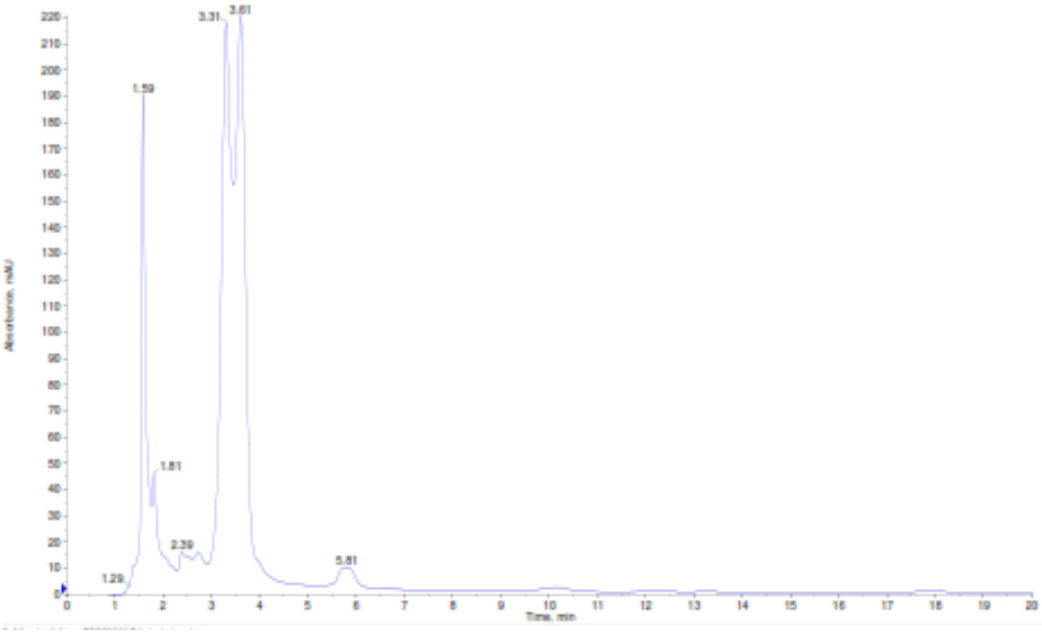

Figura 0-3. Cromatograma del Havana Club Ritual Cubano.

A.2. TIEMPOS DE RETENCIÓN HPLC (BAJO LAS CONDICIONES DE TRABAJO EMPLEADAS) PARA ALGUNOS COMPONENTES DEL RON

Compuesto	min
Fenoles y derivados	
guayacol	6,91
4-metilguayacol	10,00
(*E*)-anetol	12,90
4-vinilguayacol	13,83
siringol	15,04
eugenol	15,23
4-propilguayacol	15,37
vainillina	16,44
4-metilsiringol	17,98
cis-isoeugenol	18,00
homovainillina	18,33
1-guayacilpropino	18,74
acetovainillona	19,08
guayacilacetona	20,50
4-vinilsiringol	21,46
ácido vainillico	21,67
guayacil vinil cetona	21,92
vainillato de etilo	22,05
4-allil siringol	22,55
alcohol dihidroconiferaldehído	23,76
siringaldehído	24,15
alcohol (*Z*)-coniferílico	24,38
1-siringilpropino	24,42
(*E*)-4-propenil siringol	25,19
homosiringaldehído	25,35
homovainillato de metilo	25,82
(*E*)-coniferaldehído	26,00
acetosiringona	26,05
alcohol (*E*)-coniferílico	26,20

Compuesto	min
siringilacetona	27,13
β-hidroxipropiovainillona	28,13
siringil vinil cetona	28,56
siringato de etilo	28,74
ácido siríngico	28,95
ácido ferúlico	29,91
alcohol dihidrosinapílico	30,14
alcohol (*Z*)-sinapílico	30,72
(*E*)-sinapaldehído	32,26
alcohol (*E*)-sinapílico	32,62
dihidrovainillina	50,40
Esteres	
carbonato de dietilo	1,44
lactato de etilo	1,68
3-hidroxibutirato de etilo	3,29
acetoacetato de etilo	3,54
oxalato de dietilo	4,40
hexanoato de etilo	4,59
2-furoato de etilo	5,95
levulinato de etilo	6,24
malonato de dietilo	6,51
dietoxiacetato de etilo	7,16
benzoato de etilo	9,49
succinato de dietilo	9,81
octanoato de etilo	10,27
nicotinato de etilo	10,75
acetato de 2-feniletilo	11,69
malato de dietilo	12,51
decanoato de etilo	16,44
undecanoato de etilo	19,49

Compuesto	min
dodecanoato de etilo	22,36
citrato de etilo	24,29
tetradecanoato de etilo	27,70
hexadecanoato de etilo	32,62
linoleato de etilo	36,29
oleato de etilo	36,44
sebacato de dibutilo	36,54
octadecanoato de etilo	37,00
eicosanoato de etilo	41,19
docosanoato de etilo	45,01
tetracosanoato de etilo	48,57
Ácidos grasos	
acético	0,80
butanoico	1,46
isovalérico	1,92
valérico	2,44
hexanoico	4,28
heptanoico	6,76
2-etilhexanoico	8,02
octanoico	9,71
nonanoico	12,80
decanoico	15,90
dodecanoico	21,65
tetradecanoico	27,03
pentadecanoico	29,50
palmitoleico	31,39
hexadecanoico	32,10
oleico	35,88
octadecanoico	36,44

Compuesto	min
Alcoholes	
isobutanol	0,88
3-metil-1-butanol	1,18
2-metil-1-butanol	1,19
3-metil-2-buten-1-ol	1,39
2,2-dietoxietanol	3,18
2-etil-1-hexanol	5,37
alcohol bencílico	5,45
1-octanol	6,45
2-feniletanol	7,61
2-fenoxietanol	10,88
1-dodecanol	18,83
Terpenos	
β-pineno	3,97
mirceno	4,32
1,4-cineol	4,92
p-cimeno	5,01
limoneno	5,22
1,8-cineol	5,33
γ-terpineno	6,00
alcanfor	8,44
borneol	9,11
α-terpineol	9,91
Aldehídos y cetonas	
hexanal	1,56
2-heptanona	2,54
2(5H)-furanona	2,99
benzaldehído	3,73
nonanal	7,36
decanal	10,46
2,6-dimetoxi-1,4-benzoquinona	21,55
benzofenona	22,96
Acetales	
acetal	1,14
1,1-dietoxibutano	3,00
2-furfural dietil acetal	6,64
1,1-dietoxi-2-metilpropano	7,47
Derivados furánicos	
2-furfural	1,86
5-metil-2-furfural	3,82
5-hidroximetil-2-furfural	11,35
Lactonas	
γ-butirolactona	2,96
trans-β-metil-γ-octalactona	13,16
cis-β-metil-γ-octalactona	14,13
Otros compuestos	
2,2-dietoxietanol	3,18
No identificado 1	21,16
No identificado 2	27,85
ftalato de dibutilo	31,71
No identificado 3	34,01
bis (2-etilhexil) ftalato	44,02

A.3. CARACTERÍSTICAS DE LOS ESPECTROS UV-VIS DE LOS COMPUESTOS ORGÁNICOS

Tabla 0-1. Características de los espectros UV-Vis de loscompuestos orgánicos.

Tipo de compuesto	*Transiciones*	*Características*
Hidrocarburos saturados	$\sigma\sigma^*$	Transparentes en UV-Vis
Alquenos	$\pi\pi^*$	Etileno λ_{max}=165nm (ε=10000).Doble enlace muy sustituido o unido a auxócromos puede dar absorción final cerca de 200nm.
Polienos	$\pi\pi^*$	-Dienos: λ_{max}≈220nm (log ε 3,9-4.3) -Dienos homoanulares: λ_{max}> 250nm (log ε =3.7-3.9) (Reglas de Fieser) -Polienos: λ_{max} = 220 +30(n-2) (log ε >4)
Alquinos	$\pi\pi^*$	Absorción final cerca de 200nm
Eteres, alcoholes	$n\sigma^*$	Absorción final cerca de 200nm
Aldehidos y cetonas saturados	$n\pi^*$	λ_{max} =270-300nm ε <100
Aldehidos y cetonas $\alpha\beta$insat.	$n\pi^*$	λ_{max} =310-325nm ε <100
	$\pi\pi^*$	λ_{max} =210-235nm log ε =3.7-4.3
Acidos carbox. y derivados	$n\pi^*$	Sin absorción definida a λ > 230nm
Acidos carbox. y derivados $\alpha\beta$-insaturados	$\pi\pi^*$	λ_{max}< 230nm log ε ≈ 3.5
Aminas y aminoacidos satur.	$n\sigma^*$	Absorción final. λ_{max}< 220nm log ε ≈2
Haluros de alquilo saturados	$n\sigma^*$	Absorción final. Sólo los yoduros absorben cerca de 260nm log ε ≈ 2.5
Aromáticos (alternantes)	$\pi\pi^*$	Banda primaria: λ_{max} = 205-260nm log ε ≈4. Banda secundaria: λ_{max} = 260-300nm log ε =2.5-3.5. Conjugación puede sumergir banda secundaria. Efectos del pH sobre anilinas y fenoles.
Azometinos, oximas, nitrilos	$\pi\pi^*$	Transparentes en UV cercano, si son $\alpha\beta$insat. λ_{max}=210-230nm log ε ≈4
Azo compuestos	$n\pi^*$	λ_{max}≈350 ε < 30
Nitrocompuestos saturados	$n\pi^*$	λ_{max}≈270 ε < 30
Nitrosocompuestos	$n\pi^*$	λ_{max}≈300 ε ≈100 λ_{max}≈660 ε < 50
Sulfonas saturadas		Transparentes en UVcercano
Sulfóxidos saturados	$n\pi^*$	λ_{max}≈220 ε < 2000
Tiocarbonilos	$n\pi^*$	Posición variable usualmente λ_{max}>300nm
	$\pi\pi^*$	λ_{max}=250-320nm