

Método analítico multivariable basado en espectroscopía NIR y herramientas quimiométricas para diferenciar entre Tequila Blanco '100% agave' y 'mixto'

Multivariable analytical method based on NIR spectroscopy and chemometric tools to differentiate between '100% agave' and 'mixed' white tequila

Pérez-Beltrán, C.H.^{1*}, León-Lopez, L.^{1,2}, Jiménez-Carvelo, A.M.³, Pérez-Caballero, G.⁴, Olmos-Espejel, J.J.⁵, Cuevas-Rodríguez, E.O.^{1,2}

¹Laboratorio de Nutracéuticos, Universidad Autónoma de Sinaloa, Facultad de Ciencias Químico Biológicas, Calzada de las Américas, Nte. 2771, Ciudad Universitaria, Burócrata, 80030, Culiacán Rosales, Sinaloa, México. ²Programa de Posgrado en Ciencia y Tecnología de Alimentos, y Programa de Posgrado Integral en Biotecnología, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Culiacán, Sinaloa, México, CP 80000. ³Department of Analytical Chemistry, Faculty of Sciences, University of Granada, C/ Fuentenueva, s/n, E-18071 Granada, Spain. ⁴Unidad de Investigación Multidisciplinaria, Facultad de Estudios Superiores Cuautitlán, campo 4, Universidad Nacional Autónoma de México, Cuautitlán Izcalli, (México). ⁵Laboratorio de Análisis Instrumental y Química Analítica Verde, Facultad de Estudios Superiores Cuautitlán, UNAM.

RESUMEN

El tequila, una de las bebidas espirituosas más representativas de México, ve amenazada su autenticidad por prácticas de adulteración y falsificación, que afectan tanto a los consumidores como a la industria. La distinción entre las categorías de tequila blanco '100% agave' y 'mixto' resulta crucial para garantizar su calidad, valor cultural y comercial. En este estudio se propone un método analítico multivariable basado en espectroscopía de infrarrojo cercano (NIR) y herramientas quimiométricas para abordar este problema, mediante el uso de modelos matemáticos avanzados, incluyendo análisis de componentes principales (PCA), regresión de mínimos cuadrados parciales (PLSR), k-vecinos cercanos (kNN), análisis discriminante mediante regresión de mínimos cuadrados parciales (PLS-DA), modelado flexible e independiente por analogía de clases (SIMCA) y sistema de aprendizaje automático mediante vectores soporte (SVM). Los resultados muestran que los modelos SVM y PLS-DA lograron clasificaciones perfectas (sensibilidad, especificidad y precisión = 1), evidenciando su eficacia en la autenticación de las categorías de Tequila. Adicionalmente, el modelo de regresión SVMR demostró un desempeño sobresaliente en la predicción del contenido alcohólico, con un coeficiente de determinación (R^2) de 1.0. Este enfoque ofrece una alternativa eficiente, rápida y no destructiva frente a los métodos tradicionales de control de calidad, como la cromatografía. Su implementación en la industria tequilera y laboratorios de rutina puede mejorar significativamente la detección de fraudes y garantizar la autenticidad de los productos. Este estudio destaca el potencial de la espectroscopía NIR y la quimiometría como herramientas clave para fortalecer el control de calidad en la industria de bebidas espirituosas.

Palabras Claves: Control de calidad, Autenticación, Fraude alimentario, Quimiometría, Espectroscopía infrarrojo cercano.

ABSTRACT

Tequila, one of Mexico's most representative spirits, is threatened by adulteration and counterfeiting practices that affect both consumers and the industry. The distinction between the categories of '100% agave' and 'mixed' white Tequila is crucial to guarantee its quality, cultural and commercial value. In this study, a multivariate analytical method based on near infrared spectroscopy (NIR) and chemometric tools is proposed to address this problem, using advanced mathematical models, including principal component analysis (PCA), partial least squares regression (PLSR), k-nearest neighbors (kNN), partial least squares regression discriminant analysis (PLS-DA), soft independent modeling of class analogy (SIMCA) and support vector machine learning (SVM). Results show that the SVM and PLS-DA models achieved perfect classifications (sensitivity, specificity and precision = 1), evidencing their effectiveness in the authentication of Tequila categories. Additionally, the SVMR regression model demonstrated outstanding performance in predicting alcohol content, with a coefficient of determination (R^2) of 1.0. This approach offers an efficient, fast and non-destructive alternative to traditional quality control methods such as chromatography. Its implementation in the tequila industry and routine laboratories can significantly improve fraud detection and guarantee product authenticity. This study highlights the potential of NIR spectroscopy and chemometrics as key tools to strengthen quality control in the spirits industry.

Keywords: Quality control, Authentication, Food fraud, Chemometrics, Near infrared spectroscopy

*Autor de correspondencia: Christian Hazael Pérez Beltrán

E-mail: chperez@uas.edu.mx

ORCID ID: [0000-0003-0957-2542](https://orcid.org/0000-0003-0957-2542)

Registro ORCID Autores: LL: [0000-0003-4973-2096](https://orcid.org/0000-0003-4973-2096); JC: [0000-0002-9061-1686](https://orcid.org/0000-0002-9061-1686); OE: [0000-0001-6074-9361](https://orcid.org/0000-0001-6074-9361);

CR: [0000-0001-6648-7323](https://orcid.org/0000-0001-6648-7323); PC: [0000-0003-4567-4634](https://orcid.org/0000-0003-4567-4634)

Enviado: Diciembre 2024

Aceptado: Enero 2025

INTRODUCCIÓN

El tequila, una de las bebidas espirituosas más emblemáticas de México, es reconocido mundialmente por su calidad, autenticidad y profundo vínculo cultural con el país. Esta bebida se produce a partir de la fermentación y destilación del jugo extraído del agave azul (Agave tequilana Weber var. azul), de acuerdo con estrictas regulaciones que garantizan su autenticidad, tales como la Norma Oficial Mexicana NOM-006-SCFI-2012, vigilada por el Consejo Regulador del Tequila (CRT). Este reglamento especifica los requisitos para etiquetar un producto como "100% agave", en comparación con los tequilas mixtos, que contienen hasta un 49% de azúcares de otras fuentes (CRT, 2016). A pesar de estos controles, la adulteración y falsificación de tequila han aumentado, afectando tanto a la industria como a los consumidores. Por ello, es crucial implementar metodologías analíticas avanzadas que complementen los métodos tradicionales y faciliten una detección rápida y confiable de posibles fraudes (Warren-Vega *et al.*, 2023).

Los métodos analíticos oficiales de control de calidad alimentaria suelen estar basados en técnicas analíticas tradicionales, tales como la cromatografía de gases; usada para la determinación del contenido permitido de aldehídos, metanol y alcoholes superiores en tequilas (NMX-V-005-NORMEX-2018), o la cromatografía de líquidos; utilizada para la determinación de los niveles máximos de furfural en tequilas (NMX-V-004-NORMEX-2018).

La mayoría de las técnicas analíticas tradicionales aplican un enfoque dirigido y requieren ser llevadas a cabo en instalaciones especializadas de manera off-line con un tiempo prolongado de análisis, son de alta complejidad y, a menudo, requieren de un exhaustivo pretratamiento de la muestra. Una alternativa para disminuir estos inconvenientes y aumentar la eficiencia de los análisis de control de calidad alimentario es aplicar un enfoque no dirigido para desarrollar métodos analíticos multivariable (MAM), basados en el uso de técnicas espectroscópicas no invasivas y herramientas quimiométricas, que sean rápidos, confiables, sencillos y que el tratamiento de la muestra sea poco o nulo, para que puedan ser usados como métodos de análisis de cribado (métodos analíticos de vanguardia) para una detección más temprana de posibles fraudes alimentarios (Valcárcel y Cárdenas, 2005).

De esta manera, el uso de técnicas espectroscópicas propicia el desarrollo de estos métodos analíticos de vanguardia que permiten agilizar los análisis de control y aseguramiento de calidad de una mayor cantidad de muestras, dejando las técnicas analíticas tradicionales para el desarrollo de métodos analíticos de retaguardia (Valcárcel y Cárdenas,

2005; Jiménez-Carvelo *et al.*, 2021), como lo son los métodos cromatográficos, utilizados para confirmar ciertos resultados sospechosos de adulteración o falsificación provistos por los métodos de vanguardia.

Entre las técnicas espectroscópicas más exploradas para el control de calidad en la industria alimentaria se encuentran la espectroscopía de infrarrojo medio con transformada de Fourier (FTIR), de infrarrojo cercano (NIR) y/o Raman (Arendse *et al.*, 2021; Tirado-Kulieva *et al.*, 2022), cuyos espectros son habitualmente utilizados para desarrollar de manera individual métodos analíticos multivariable.

En este contexto, la quimiometría ha emergido como una herramienta esencial en el análisis de alimentos y bebidas, permitiendo interpretar grandes volúmenes de datos generados por técnicas instrumentales modernas. Definida como la ciencia que aplica métodos matemáticos y estadísticos para maximizar la extracción de información química (Hibbert, 2016), la quimiometría combina modelos supervisados y no supervisados para clasificar, predecir y evaluar patrones químicos. Su integración con técnicas espectroscópicas, como la espectroscopía de infrarrojo cercano (NIR), representa una solución innovadora para superar las limitaciones de los métodos analíticos tradicionales, al proporcionar análisis rápidos, no destructivos y más económicos (Arendse *et al.*, 2021).

La espectroscopía NIR se ha consolidado como una técnica versátil en la industria alimentaria y de bebidas, utilizada para evaluar parámetros de calidad, autenticidad y composición química (Pontes *et al.*, 2006; Keong Tan *et al.*, 2023). En ese sentido, también se han empleado diversas técnicas espectroscópicas en la industria tequilera, tales como la ultravioleta visible (UV-Vis) (Contreras *et al.*, 2010), espectroscopía infrarroja con transformada de Fourier (FTIR) (Pérez-Beltrán *et al.*, 2021) o espectroscopía Raman compensada espacialmente (Pérez-Beltrán *et al.*, 2022).

Cuando se combina con herramientas quimiométricas, como regresión de mínimos cuadrados parciales (PLSR), análisis discriminante mediante regresión PLS (PLS-DA) o máquinas de soporte vectorial (SVM), es posible diferenciar entre productos auténticos y adulterados con una alta precisión (Kolomiets *et al.*, 2010; Fernández-Lozano *et al.*, 2019; Warren-Vega *et al.*, 2023).

El presente trabajo se enfoca en desarrollar y validar un método analítico multivariable basado en espectroscopía NIR y quimiometría para diferenciar entre tequila blanco '100% agave' y 'mixto', así como para predecir su contenido alcohólico. Este enfoque busca proporcionar una

herramienta robusta y eficiente para el control de calidad en la industria tequilera, contribuyendo a la preservación de la autenticidad y calidad de esta bebida icónica.

MATERIALES Y MÉTODOS

MUESTRAS

Un total de 55 muestras de Tequila Blanco de distintas marcas fueron empleadas en este estudio, de las cuales 30 pertenecían a la categoría '100% agave' (TB - Tequila Blanco) y 25 a la categoría 'mixto' (TBM - Tequila Blanco Mixto). Dichas muestras fueron otorgadas por el Consejo Regulador del Tequila de México (CRT, 2016).

Instrumentación

Se utilizó un espectrómetro Frontier IR Dual-Range System, PerkinElmer, equipado con una esfera de integración NIR IntegratIRTM de 50.8 mm (2") de diámetro, PIKE Technologies. Se analizaron 2 mL de cada muestra de Tequila Blanco, sin ningún tipo de tratamiento, los cuales se vertieron en una celda de cuarzo y fueron cubiertos con un émbolo de oro de 1.0 mm de paso óptico para evitar la dispersión de la radiación electromagnética. Las huellas instrumentales NIR se obtuvieron en el rango 10000-4000 cm^{-1} , usando 16 escaneos, con una resolución nominal de 4 cm^{-1} y correcciones de fondo entre el análisis de una muestra y otra. Las huellas instrumentales obtenidas fueron transformadas de transmitancia a absorbancia y corregidas manualmente en su línea base en los puntos 10000, 9400.67, 8868.69, 7784.51, 6006.13, 5461.28 y 4000 cm^{-1} , mediante el software SpectrumTM 10 STD, PerkinElmer. Finalmente, se digitalizaron y exportaron en formato ASC (Action Script).

ANÁLISIS DE DATOS MULTIVARIABLE

Las huellas instrumentales originales NIR, formadas por 6000 variables cada una (1 dato por cm^{-1}), fueron exportadas a formato MATLAB (Mathworks, Massachusetts, USA, v. R2017b). El conjunto de datos fue dividido en un subconjunto de entrenamiento conformado por 42 muestras (24 TB / 18 TBM) y un subconjunto de validación externa formado por 11 muestras (6 TB / 5 TBM). Dicha selección de muestras se realizó de manera aleatorizada en una proporción 80-20% para cada uno de los subconjuntos previamente mencionados.

Los análisis de datos multivariable fueron llevados a cabo con el software PLS_Toolbox (v. 8.6.1, 2019, Eigenvector Research In., Manson, WA, USA).

Las herramientas quimiométricas empleadas para el análisis exploratorio fueron análisis de componentes principales (PCA, Principal Component Analysis) y regresión parcial lineal mediante mínimos cuadrados (PLSR, Partial Least Square Regression), mientras que para los estudios de clasificación se emplearon k-vecinos cercanos (kNN, k-Nearest Neighbor), modelado flexible e independiente por analogía de clases (SIMCA, Soft Independent Modeling of Class Analogy), análisis discriminante mediante regresión parcial de mínimos cuadrados (PLS-DA, PLS-Discriminant Analysis) y sistema de aprendizaje automático mediante vectores soporte (SVMs, Support Vector Machines). Para la predicción del contenido alcohólico se utilizó PLSR y regresión mediante SVMR. Asimismo, para los análisis exploratorios se aplicaron conjuntamente los siguientes pre-procesados: filtro Whitaker y 1ª derivada; para los estudios de clasificación se seleccionó el rango 6000-4000 cm^{-1} , se normalizó (área = 1) y se realizó un centrado en la media; mientras que para la predicción del contenido alcohólico se utilizaron las 6000 variables a las cuales se les aplicó el filtro Whitaker y 1ª derivada como pre-procesado.

RESULTADOS Y DISCUSIÓN

El Tequila contiene una gran variedad de compuestos volátiles (Soto-Romero *et al.*, 2016), como alcoholes superiores, aldehídos, ácidos grasos, ésteres, compuestos azufrados, algunos compuestos fenólicos y hasta antioxidantes, los cuales le otorgan su aroma y sabor, éstos han sido identificados a través de cromatografía de gases (Vallejo-Córdoba *et al.*, 2004; Peña-Alvares *et al.*, 2006) y de líquidos (Bukovsky-Reyes *et al.*, 2018), entre otros métodos analíticos (Warren-Vega *et al.*, 2023). No obstante, el Tequila está constituido mayoritariamente por agua y etanol en proporciones aproximadas de 60% y 40% (NOM-006-SCFI-2012), respectivamente.

La presencia de ambos componentes mayoritarios se puede corroborar en la Figura 1, donde se muestran las vibraciones características de agua y etanol de dos huellas instrumentales de Tequila Blanco, categorías TB (espectro color rojo) y TBM (espectro color verde), obtenidas mediante espectroscopía NIR.

Las intensas bandas alrededor de 7000 cm^{-1} y 5200 cm^{-1} se atribuyen a las bandas características de combinación de $\nu_1 + \nu_3$ y $\nu_1 + \nu_3$, respectivamente, del agua, las cuales están superpuestas con las bandas de combinación del etanol $2\nu(\text{OH})$ y $\nu(\text{OH}) + \delta(\text{OH})$; mientras que las regiones 6000-5500 cm^{-1} e inferior a 4500 cm^{-1} a las vibraciones

$2\times\nu(\text{CH})$ y $\nu(\text{CH}) + \delta(\text{CH})$ del etanol (Livermore *et al.*, 2003; Pontes *et al.*, 2006; Kolomiets *et al.*, 2010).

Tal como se puede apreciar en la Figura 1, las señales instrumentales de ambas categorías de tequila son muy similares, lo que imposibilita su diferenciación y autentificación a simple vista. Es por ello, que se requiere la aplicación de herramientas quimiométricas que permitan extraer información relevante de dichas señales para lograr su adecuada autentificación, tal como se detalla en los siguientes subapartados.

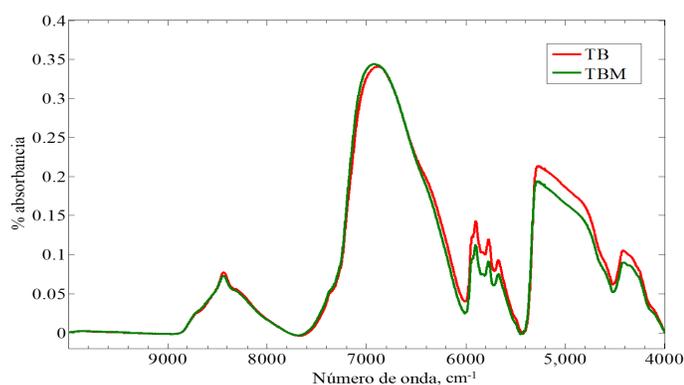


Figura 1. Espectros NIR de Tequila Blanco. TB: '100% agave' y TBM: 'mixto'.

ANÁLISIS EXPLORATORIO

El análisis exploratorio se llevó a cabo para estudiar el agrupamiento y/o comportamiento natural de este conjunto de datos obtenido de muestras de Tequila Blanco de categorías '100% agave' y 'mixto', y para detectar posibles muestras anómalas. Se realizaron diversos pre-procesados, encontrando que los mejores resultados para estos análisis exploratorios eran obtenidos al aplicar conjuntamente el filtro Whitaker y 1ª derivada.

El primer análisis exploratorio fue llevado a cabo mediante PCA, el cual se construyó con 6 componentes principales (PCs, principal components), mismas que explicaban el 99.2% de la varianza total. Al realizar el análisis de la PC1 vs PC2 y graficar sus puntuaciones (véase Material Suplementario [Figura 1 \(a\)](#)), se observó que las muestras se ordenaban en el plano según su contenido alcohólico, siendo las de menor graduación alcohólica las situadas en la parte inferior y las de mayor graduación alcohólica las situadas en la parte superior. Asimismo, dicha gráfica de puntuaciones propició la detección de dos muestras anómalas (círculos marcados en magenta), las cuales presentaban una señal instrumental distinta en el intervalo 6000-5500 cm^{-1} ([Figura 1 \(b\)](#)),

debido a que el análisis espectroscópico de dichas muestras se realizó con una cantidad inferior a la requerida, por lo que dichas señales instrumentales fueron excluidas de los siguientes análisis de datos multivariable. La utilización de la herramienta quimiométrica de PCA también fue utilizada por Contreras *et al.* (2010) para desarrollar un modelo con 2 PCs (99.99% de la varianza total), con el cual se obtuvo una clara agrupación de Tequilas Blanco y Tequilas Reposado, lo cual se debió a las características inherentes del proceso de añejamiento sucedido en las barricas de madera, facilitando así dichas agrupaciones. Posteriormente, se realizó un segundo análisis exploratorio mediante PLSR, el cual fue construido con 6 variables latentes (LVs, latent variables), que explicaban el 99.1% y 92.7% de la varianza total en los bloques X e Y, respectivamente, y cuyo gráfico de puntuaciones (LV1 vs LV2) se observa en la Figura 2. En este caso y a diferencia del gráfico de puntuaciones del PCA, se pueden distinguir claramente las agrupaciones de las muestras pertenecientes a ambas categorías de tequila, lo cual demuestra la ventaja del PLSR sobre el PCA debido a su capacidad para capturar tanto la varianza y correlación entre los datos (Wise *et al.*, 2006).

Una vez finalizado el análisis exploratorio, dichos datos fueron empleados en los siguientes análisis de datos multivariable. Dichos resultados son consistentes con otras investigaciones realizadas previamente, aún cuando la técnica espectroscópica fue diferente; por un lado, se utilizó espectroscopía FTIR con fusión de datos (Pérez-Beltrán *et al.*, 2021) y, por el otro, espectroscopía Raman compensada espacialmente (SORS, Spatially Offset Raman Spectroscopy) (Pérez-Beltrán *et al.*, 2022). En ambos estudios la herramienta quimiométrica que mejores agrupaciones mostró fue la de PLS, evidenciando tendencias claras para las categorías de "100% agave" y "mixto" de las muestras de Tequila Blanco.

ANÁLISIS DE CLASIFICACIÓN

Para la realización de este tipo análisis se emplearon distintas herramientas quimiométricas de análisis supervisado, en las cuales se consideró a la categoría '100% agave' (TB) como clase diana, debido a que es la categoría con mayor valor económico y, por tanto, más propensa a sufrir adulteraciones y/o falsificaciones. Después de un estudio exhaustivo de distintos pre-procesados, se seleccionaron únicamente 2000 variables (6000-4000 cm^{-1}) de las 6000 variables originales para eliminar la mayor cantidad posible de interferencia debida a las características bandas de combinación ($\nu_1 + \nu_3$) del agua.

Seguidamente, dichos datos fueron normalizados (área = 1) y centrados en la media. A continuación, se describen los resultados obtenidos para cada uno de los modelos matemáticos.

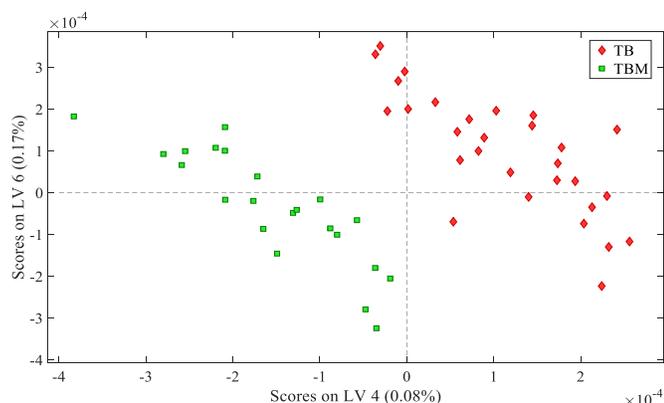


Figura 2. Gráfico de puntuaciones (LV4 vs LV6) de análisis exploratorio mediante PLSR a conjunto de datos de Tequila Blanco '100% agave' (TB) y Tequila Blanco 'Mixto' (TBM).

SVM

Para la elaboración de este modelo matemático se utilizó la función Kernel de 'base radial' con valores 'gama' y 'costo' comprendidos entre los rangos 10^{-6} - 10 y 10^{-3} - 10^2 , respectivamente; así como una compresión adicional mediante PLS con 16 LVs. Se estableció un valor de 0.5 como límite de decisión para la clasificación de las muestras, donde un valor > 0.5 correspondía a las muestras pertenecientes a la clase TB, mientras que valores < 0.5 a muestras pertenecientes a la clase TBM. La gráfica de clasificación, representada por la Figura 3 (a), muestra los resultados de los conjuntos de entrenamiento y validación externa en donde todas las muestras fueron correctamente diferenciadas en sus clases correspondientes. Los resultados de este conjunto de validación se ponen de manifiesto en la tabla de contingencias de validación, expuesta a través de la Figura 3 (b). A diferencia de esta investigación, Contreras *et al.* (2010) utilizaron espectroscopía UV-Vis en conjunto con SVM y PCA para diferenciar entre Tequilas Blancos y Tequilas reposados, obteniendo una clasificación 100% correcta. Por su parte, en el estudio desarrollado por Pérez-Beltrán *et al.* (2022), se utilizó espectroscopía Raman compensada espacialmente y SVM, en donde la diferenciación de las categorías "100% agave" y "mixto" de Tequila Blanco fue 100% correcta, logrando resultados consistentes con esta investigación.

PLS-DA

El modelo PLS-DA fue construido utilizando 18 LVs, las cuales explicaban el 100% y 98.3% de la varianza acumulada en los bloques X e Y, respectivamente, con un error promedio de clasificación de validación cruzada (RMSECV, Root Mean Square Error for Cross Validation) de 0.604. Al igual que en el modelo SVM, se estableció un límite de 0.5 como criterio de decisión, asociando los valores > 0.5 a muestras de la clase TB y valores < 0.5 a muestras de la clase TBM, tal como se muestra en su gráfica de clasificación, representada por la Figura 4 (a). Asimismo, en ella se aprecia la correcta clasificación de todas las muestras de los conjuntos de entrenamiento y validación externa, cuyos resultados de validación se recopilan en su correspondiente tabla de contingencias en la Figura 4 (b). Dichos resultados, coinciden con los obtenidos por Pérez-Beltrán *et al.* (2021, 2022) en donde el modelo de clasificación de PLS-DA fue capaz de diferenciar correctamente todas las muestras pertenecientes a las dos categorías diferente de Tequila Blanco, evidenciando el potencial que tiene esta herramienta para obtener resultados consistentes independientemente de la técnica espectroscópica utilizada.

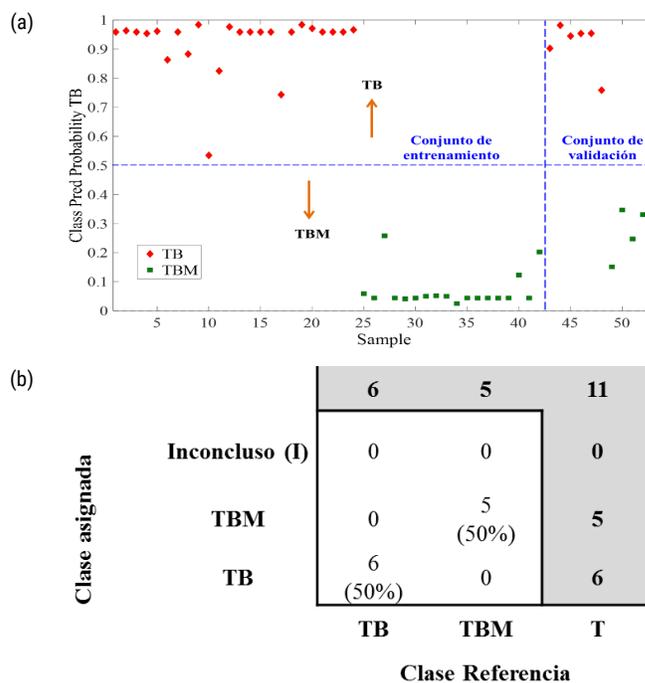


Figura 3. (a) Gráfica de clasificación y (b) tabla de contingencias de validación para el modelo de clasificación SVM. Clase diána: TB-Tequila Blanco '100% agave'; Clase alternativa: TBM-Tequila Blanco 'Mixto'.

kNN

Para la construcción de este modelo matemático se consideró como número óptimo de vecino un valor de 8 ($k = 8$), con el cual fue posible encontrar una pequeña diferenciación entre ambas clases bajo estudio. El límite de decisión se mantuvo igual que en los modelos SVM y PLS-DA, siendo los valores > 0.5 asociados a las muestras de la clase TB y valores < 0.5 a la clase TBM, tal como se muestra en la gráfica de clasificación del Material Suplementario [Figura 2 \(a\)](#).

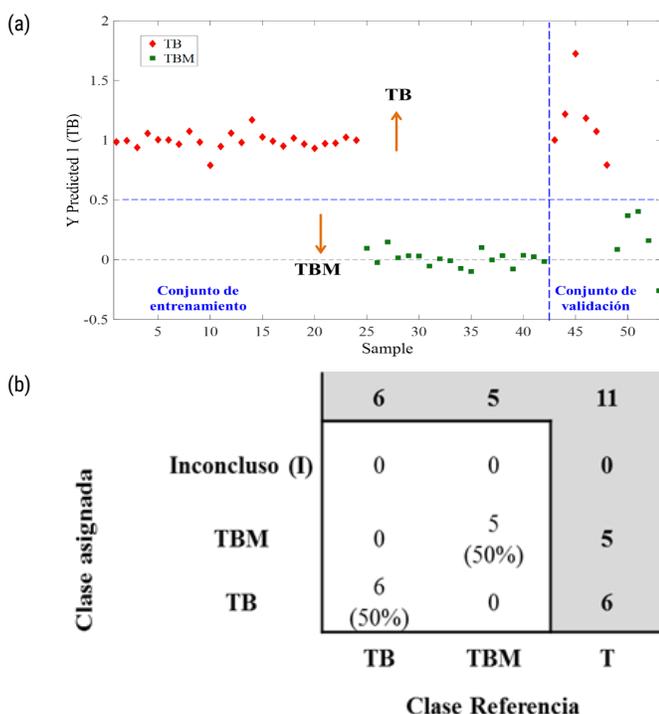


Figura 4. (a) Gráfica de clasificación y (b) tabla de contingencias de validación para el modelo de clasificación PLS-DA. Clase diana: TB-Tequila Blanco '100% agave'; Clase alternativa: TBM-Tequila Blanco 'Mixto'.

En este caso, el modelo matemático presentó dificultades para diferenciar las muestras del conjunto de entrenamiento, lo cual se vio reflejado en la validación del modelo con 3 muestras mal clasificadas (círculos marcados en color magenta). Los resultados que pueden apreciarse en la tabla de contingencias de validación en el [Material Suplementario](#).

SIMCA

Para encontrar los mejores resultados se estudiaron dos estrategias en la construcción del modelo matemático SIMCA: (i) modelo de

clasificación con dos clases de entrada (2iC-SIMCA), en el cual se utilizaron tanto la clase diana (TB) y la clase alternativa (TBM) para entrenar el modelo, y (ii) modelo de clasificación con una clase de entrada (1iC-SIMCA), en el cual sólo la clase diana fue utilizada para la etapa de entrenamiento del modelo.

A pesar del gran potencial de esta herramienta quimiométrica y de las dos estrategias estudiadas, no fue posible obtener una clara diferenciación de ambas clases de tequila, tal como se aprecia en la gráfica de clasificación del Material Suplementario [Figura 3 \(a\)](#), en donde se muestran los estadísticos T^2 de Hotelling y Q para la clase diana con un nivel de confianza del 95%. Las muestras con valores de T^2 y $Q < 1$ son asociadas a la clase TB, mientras que las muestras con valores de T^2 y $Q > 1$ a la clase TBM. Los resultados para el conjunto de validación se pueden apreciar en la tabla de contingencias en el [Material Suplementario](#).

De manera complementaria, se calcularon las cuatro 4 principales métricas de calidad en el desempeño de cada uno de los modelos de clasificación mencionados –sensibilidad (SENS), especificidad (ESPEC), valor predictivo positivo o precisión (PREC) y el valor predictivo negativo (VPN) (Cuadros-Rodríguez *et al.*, 2016) – las cuales se muestran en la Tabla 1. Los mejores resultados, con valores = 1, se obtuvieron con los modelos matemáticos SVM y PLS-DA, seguidos del modelo kNN y por último 1iC-SIMCA.

Tabla 1. Principales métricas de calidad en el desempeño para los modelos de clasificación SVM, PLS-DA, kNN y 1iC-SIMCA.

Métricas	SVM	PLS-DA	kNN	1iC-SIMCA
	Clase diana (TB, Tequila Blanco 100% agave)			
Sensibilidad (SENS)	1.00	1.00	0.83	0.17
Especificidad (ESPEC)	1.00	1.00	0.60	0.40
Valor predictivo positivo o precisión (PREC)	1.00	1.00	0.83	0.33
Valor predictivo negativo (VPN)	1.00	1.00	1.00	0.33

En este sentido, los modelos SVM y PLS-DA son capaces de clasificar correctamente todas las muestras TB y TBM, lo cual se demuestra a través de las métricas SENS y ESPEC = 1, respectivamente. Asimismo, ambos modelos poseen una excelente precisión (PREC = 1), lo cual indica la proporción de muestras TB bien clasificadas con relación a todas las muestras totales asignadas a la clase TB; así como también un excelente VPN = 1, lo cual indica la habilidad de ambos modelos para proveer la proporción de muestras TBM bien clasificadas con relación a las muestras totales asignadas a la clase TBM.

Cabe mencionar que, a pesar de que el modelo kNN clasificó mal solamente una muestra de TB, obteniendo buenos resultados de SENS y

PREC = 0.83 en ambas métricas de calidad, sus resultados de ESPEC = 0.60 indican su falta de habilidad para clasificar muestras de la clase TBM. Por tanto, los modelos matemáticos SVM y PLS-DA arrojan mayor seguridad en los resultados y, por tanto, para su implementación en análisis de rutina y cribado para el control de calidad del Tequila Blanco, ya que son capaces de clasificar ambas categorías correctamente y de "aprender" de los datos que se van incluyendo para aumentar el conjunto de entrenamiento, por lo que la clasificación de nuevas muestras desconocidas será cada vez más confiable. Del mismo modo, se puede concluir que estas dos herramientas quimiométricas son altamente eficaces para la diferenciación de las dos categorías de Tequila Blanco ("100% agave" y "mixto"), lo cual resulta más complicado debido a que las diferencias químicas entre sí son más sutiles, debiéndose solamente al tipo de azúcar utilizado en la fermentación para su elaboración.

Predicción del contenido alcohólico.

de regresión PLSR y SVMR para predecir su contenido alcohólico. El primero de ellos se construyó con 6 LVs, las cuales explicaban el 99.24% y 96.90% de la varianza acumulada en los bloques X e Y, respectivamente, con un RMSEC de 1.481 y $R^2 = 0.969$ (Figura 5 (a)). El segundo modelo de regresión, SVMR, fue construido utilizando 22 vectores soporte (SVs, support vectors), función Kernel de base radial, y valores óptimos de gama y costo de 0.00032 y 100, respectivamente. Dicho modelo matemático proporcionó un RMSEC de 0.091 y un $R^2 = 1.0$ (Figura 5 (b)). Además de las métricas de calidad en el desempeño de la predicción del contenido alcohólico comentadas anteriormente, se calcularon otras 5 métricas, sugeridas en las prácticas estandarizadas para la validación de calibraciones multivariable realizadas empíricamente (ASTM E2617, 2017).

Como se puede observar a partir de las métricas de calidad de la Tabla 2, el modelo de regresión SVMR obtuvo mejores resultados que el modelo PLSR en cuanto a los errores que estos modelos pueden presentar al momento de realizar las predicciones. No obstante, el porcentaje de la desviación estándar de los residuos de validación (SDV, Standard Deviation of Validation residuals), el cual hace referencia a la concordancia entre el valor predicho y el valor de referencia, de ambos modelos es muy similar, indicando que las predicciones del contenido alcohólico de ambos modelos son muy próximas entre sí.

Tabla 2. Métricas de calidad en el desempeño de la predicción del contenido alcohólico de muestras de Tequila Blanco mediante SVMR y PLSR.

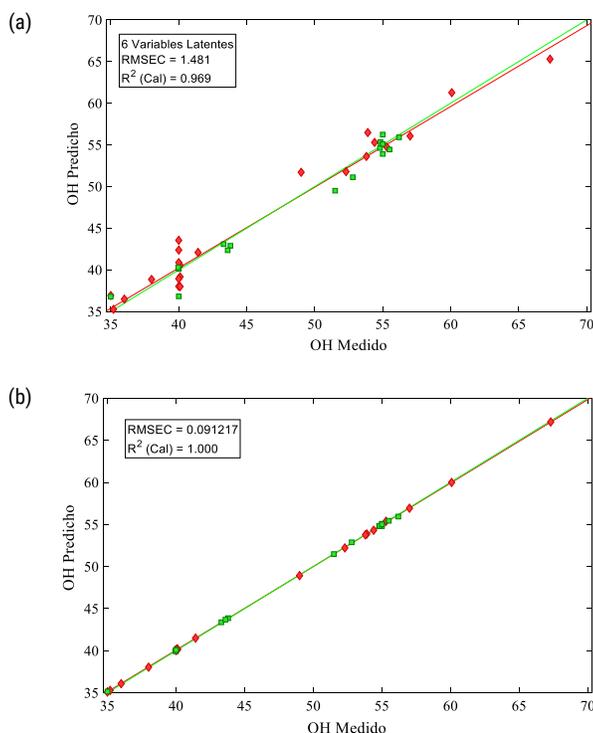


Figura 5. Curvas de calibración para la cuantificación del contenido alcohólico de muestras de Tequila Blanco mediante los modelos matemáticos (a) PLSR y (b) SVMR

El conjunto de entrenamiento estuvo conformado por 38 muestras (22 TB / 16 TBM) y el conjunto de validación externa por 9 muestras (6 TB / 3 TBM), los cuales se utilizaron para construir los modelos matemáticos

Métricas	PLSR	SVMR
	Valor (%)	
Coefficiente de determinación (R^2)	0.969	1.00
Error cuadrático medio (RMSE)	7.21	6.22
Error absoluto medio (MAE)	5.17	3.79
Error absoluto de la mediana (MdAE)	6.16	5.02
Error estándar de validación (SEV)	6.86	6.47
Desviación estándar de los residuos de validación (SDV)	6.47	6.21

CONCLUSIONES

La adulteración y falsificación de bebidas alcohólicas espirituosas, como el tequila, ha ido en aumento a través de los años, con lo cual se requieren de métodos analíticos auxiliares para agilizar y maximizar la detección de dichas bebidas apócrifas. En este trabajo, se ha desarrollado un método analítico multivariable basado en la espectroscopía de infrarrojo cercano (NIR) y herramientas quimiométricas para el control de calidad de las categorías '100% agave' y 'mixto' del Tequila Blanco. Las

herramientas quimiométricas con las cuales fue posible obtener información de relevancia de las huellas instrumentales y, por tanto, mejor desempeño fueron SVM y PLS-DA, con las cuales ha sido posible diferenciar entre ambas categorías de Tequila Blanco, logrando una sensibilidad, especificidad, precisión y valor predictivo negativo = 1. Del mismo modo, se desarrolló un modelo matemático para predecir el contenido alcohólico de las muestras, con el cual fue posible obtener un $R^2 = 1.0$ al utilizar regresión mediante SVMR.

Dichos métodos analíticos multivariados podrían ser implementados en laboratorios analíticos de rutina para agilizar el análisis de muestras y aumentar la detección de bebidas alcohólicas adulteradas o falsificadas, así como también podría ser utilizado en la industria tequilera bajo condiciones reales de producción para agilizar el control de calidad de los productos elaborados en sus instalaciones, ayudando a garantizar su aplicabilidad práctica y confiabilidad en escenarios cotidianos. Este enfoque puede extenderse a la autenticación y control de calidad de otras bebidas alcohólicas, como mezcal, sotol y whisky, enfrentando desafíos similares de adulteración y falsificación, lo cual impacta de manera simultánea en la actualización de normativas nacionales e internacionales, promoviendo estándares modernos de autenticación que incluyan técnicas rápidas, no invasivas y ambientalmente sostenibles.

AGRADECIMIENTOS

A.M.J.C. agradece la beca (RYC2021-031993-I) otorgada por MCIN/AEI/501100011033 y "European Union NextGenerationEU/PRTR".

DECLARACIÓN DE CONFLICTOS DE INTERÉS

Los autores de este trabajo declaran no tener conflictos de intereses.

REFERENCIAS

Arendse, E., Nieuwoudt, H., Magwaza, L. S., Nturambirwe, J. F. I., Fawole, O. A. and Opara, U. L. (2021). Recent advancements on vibrational spectroscopic techniques for the detection of authenticity and adulteration in horticultural products with specific focus on oils, juices and powders. *Food and Bioprocess Technology* 14, 1-22. <https://doi.org/10.1007/s11947-020-02505-x>

ASTM E2617-17. (2017). Standard practice for validation of empirically derived multivariate calibrations. ASTM International.

Bukovsky-Reyes, S. E. R., Lowe, L. E., Brandon, W. M. and Owens, J. E. (2018). Measurement of antioxidants in distilled spirits by a silver nanoparticle assay. *Journal of the Institute of Brewing* 124, 291-299. <https://doi.org/10.1002/jib.496>

Consejo Regulador del Tequila, CRT. (2016). Disponible en: <https://crt.org.mx/>. Acceso: Diciembre 12, 2024.

Contreras, U., Barbosa-García, O., Pichardo-Molina, J. L., Ramos-Ortiz, G., Maldonado, J. L., Meneses-Nava, M. A., Ornelas-Soto, N. E., López-de-Alba, P. L. (2010). Screening method for identification of adulterate and fake tequilas by using UV-VIS spectroscopy and chemometrics. *Food Research International* 43, 2356-2362. <https://doi.org/10.1016/j.foodres.2010.09.001>

Cuadros-Rodríguez, L., Pérez Castaño, E. and Ruiz Samblás, C. (2016). Quality performance metrics in multivariate classification methods for qualitative analysis. *Trends in Analytical Chemistry* 80, 612-624. <https://doi.org/10.1016/j.trac.2016.04.02>

Fernández-Lozano, C., Gestal-Pose, M., Pérez-Caballero, G., Revilla-Vázquez, A. L., and Andrade-Garda, J. M. (2019). Multivariate classification techniques to authenticate Mexican commercial spirits. In: *Quality control in the beverage industry*, (A. M. Grumezescu & A. M. Holban, eds.), Pp. 259-288. Woodhead Publishing, Elsevier.

Hibert, D. B. (2016). Introduction to multivariate data analysis and chemometrics. *Analytical and Bioanalytical Chemistry* 408(9), 2053-2054.

Jiménez-Carvelo, A. M., Martín-Torres, S., Cuadros-Rodríguez, L., González-Casado, A. (2021). Nontargeted fingerprinting approaches. In *Food authentication and traceability*, (C.M. Galankis, ed.), Pp. 163-193. Academic Press. <https://doi.org/10.1016/B978-0-12-821104-5.00010-6>

Keong Tan, W., Husin, Z., Luqman Yasruddin, M., Hakim Ismail, M. A. (2023). Recent technology for food and beverage quality assessment: a review. *Journal of Food Science and Technology* 60, 1681-1694. <https://doi.org/10.1007/s13197-022-05439-8>

- Kolomiets, O. A., Lachenmeier, D. W., Hoffmann, U. and Siesler, H. W. (2010) Quantitative determination of quality parameters and authentication of vodka using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* 18, 59-67. <https://doi.org/10.1255/jnirs.866>
- Livermore, D., Wang, Q. and Jackson, R. S. (2003). Understanding near infrared spectroscopy and its applications in the distillery. In *The alcohol textbook, a reference for the beverage, fuel and industrial alcohol industries*, (K.A. Jacques, T.P. Lyons and D.R. Kelsall, eds.), Pp. 145-170. Nottingham University Press. Disponible en: https://edisciplinas.usp.br/pluginfile.php/2993570/mod_resource/content/1/The_Alcohol_Textbook-%204%20Ed.pdf Acceso: Diciembre 12, 2024.
- Norma Mexicana NMX-V-004-NORMEX-2018. Bebidas alcohólicas- Determinación de furfural- Métodos de ensayo (prueba). Organismo Nacional de Normalización Sociedad Mexicana de Normalización y Certificación, S.C. Comité Técnico de Normalización Nacional para Bebidas Alcohólicas.
- Norma Mexicana NMX-V-005-NORMEX-2018. Bebidas alcohólicas- Determinación de aldehídos, ésteres, metanol, y alcoholes superiores-Métodos en ensayo (prueba). Organismo Nacional de Normalización Sociedad Mexicana de Normalización y Certificación, S.C. Comité Técnico de Normalización Nacional para Bebidas Alcohólicas.
- Norma Oficial Mexicana NOM-006-SCFI-2012, Bebidas alcohólicas- Tequila-Especificaciones. Comité Consultivo Nacional de Normalización de Seguridad al Usuario, Información Comercial y Prácticas de Comercio (CCNNSUICPC).
- Pérez-Beltrán, C. H., Pérez-Caballero, G., Andrade, J. M., Cuadros-Rodríguez, L., Jiménez-Carvelo, A. M. (2022). Non-targeted spatially offset Raman spectroscopy-based vanguard analytical method to authenticate spirits: White Tequilas as a case study. *Microchemical Journal* 183, 108126. <https://doi.org/10.1016/j.microc.2022.108126>
- Pérez-Beltrán, C. H., Zúñiga-Arroyo, V. M., Andrade, J. M., Cuadros-Rodríguez, L., Pérez-Caballero, G., Jiménez-Carvelo, A. M. (2021). A sensor-based methodology to differentiate pure and mixed White Tequilas based on fused infrared spectra and multivariate data treatment. *Chemosensors* 9, 1-13. <https://doi.org/10.3390/chemosensors9030047>
- Peña-Alvares, A., Capella, S., Juárez, R. and Labastida, C. (2006). Determination of terpenes in tequila by solid phase microextraction-gas chromatography-mass spectrometry. *Journal of Chromatography* 1134, 291-297. <https://doi.org/10.1016/j.chroma.2006.09.005>
- Pontes, M., Santos, S., Araujo, M., Almeida, L., Lima, R., Gaião, E. and Souto, U. (2006). Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. *Food Research International* 39, 182-189. <https://doi.org/10.1016/j.foodres.2005.07.005>
- Soto-Romero, L., Gutiérrez-Osnaya, L. J. and Trejo-Fragoso, L. (2016) Revisión de los compuestos responsables del olor y sabor del Tequila. *Investigación y Desarrollo en Ciencia y Tecnología de Alimentos* 1, 910-915. <http://www.fcb.uanl.mx/IDCyTA/files/volume1/1/10/157.pdf>
- Tirado-Kulieva, V. A., Hernandez-Martinez, E. and Suomela, J. P. (2022). Non-destructive assessment of vitamin C in food: a review of the main findings and limitations of vibrational spectroscopic techniques. *European Food Research and Technology* 248, 2185 – 2195. <https://doi.org/10.1007/s00217-022-04023-w>
- Valcárcel, M. and Cárdenas, S. (2005). Vanguard-rearguard analytical strategies. *Trends in Analytical Chemistry* 24, 67–74. <https://doi.org/10.1016/j.trac.2004.07.016>
- Vallejo-Córdoba, B., González-Córdova, A. F. and Estrada-Montoya, M. C. (2004). Tequila volatile characterization and ethyl ester determination by solid phase microextraction gas chromatography/mass spectrometry analysis. *Journal of Agricultural and Food Chemistry* 52, 5567-5571. <https://doi.org/10.1021/jf0499119>
- Warren-Vega, W. M., Fonseca-Aguñaga, R., González-Gutiérrez, L. V. and Romero-Cano, L. A. (2023). A critical review on the assessment of the quality and authenticity of tequila by different analytical techniques: Recent advances and perspectives. *Food Chemistry* 408, 135223. <https://doi.org/10.1016/j.foodchem.2022.135223>

Wise, B. M., Gallagher, N. B., Bro, R., Shaver, J. M., Winding, W. and Koch, R. S. (2006). Chemometrics Tutorial for PLS_Toolbox and Solo. Eigenvector Research, Inc. Wenatchee, WA, USA.