



UNIVERSIDAD NACIONAL DEL NORDESTE

FACULTAD DE CIENCIAS EXACTAS Y NATURALES Y AGRIMENSURA

**“ANÁLISIS MULTIELEMENTAL Y
QUIMIOMÉTRICO CLASIFICATORIO DE UVAS Y
PRODUCTOS DERIVADOS PARA EVALUAR SU
ORIGEN GEOGRÁFICO Y TRAZABILIDAD”**

Lic. Brenda Vanina Canizo

Director: Dr. Rodolfo G. Wuilloud

Co-Director: Dr. Roberto G. Pellerano

Tesis para optar al grado académico de
Doctora de la Universidad del Nordeste en Química

2020

AGRADECIMIENTOS

Durante los cinco años de estudio y trabajo del doctorado, fueron muchas las personas que me apoyaron de una u otra manera y alentaron a seguir adelante, por lo que me gustaría tener un recuerdo de agradecimiento hacia todas ellas.

En primer lugar agradezco y dedico esta tesis a mis hijos, Hana y Franco, a mi compañero de vida Patricio y a mi querido Padre, pilares presentes en esta etapa de crecimiento profesional y personal, quienes siempre me brindaron su apoyo y me acompañaron en estos cinco años de trabajo del doctorado.

A mis hermanos, quienes me brindaron su aliento y ánimo, especialmente a mi hermana Romina quién siempre estuvo para alentarme para seguir por este camino, el cual compartimos.

A mis amigos de la vida cosechados en distintas etapas de ella.

A todos mis compañeros de grupo del laboratorio por su apoyo, colaboración y amistad en estos años del doctorado.

A Ana Laura Diedrichs de R-Ladies, por ayudarme, guiarme y enseñarme sobre el uso del software "R", que fue una pieza clave en el desarrollo de esta tesis.

El presente trabajo de tesis doctoral fue realizado en el Laboratorio de Química Analítica para Investigación y Desarrollo (QUIANID) del Instituto Interdisciplinario de Ciencias Básicas (ICB-CONICET UNCUYO) y la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Cuyo. Por lo que deseo expresar un gran agradecimiento a dichas instituciones y a mi director de tesis, el Dr. Rodolfo Wuilloud, quien me dio la posibilidad de crecer profesionalmente y guio en estos años de trabajo, estando siempre presente para facilitar y animarme a seguir para la conclusión de esta tesis doctoral. Agradezco a mi co-director, el Dr. Roberto Pellerano, quién a la distancia fue un pilar importante en la concreción de todos los trabajos realizados, brindándome sus conocimientos, apoyo y comprensión. Agradezco enormemente la paciencia de ambos.

También deseo expresar mi agradecimiento a la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste, especialmente a la Srta. Liz Barberán quién atendió a todas mis consultas a la distancia.

Finalmente agradezco al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), por haber financiado mediante una beca mis estudios doctorales permitiendo así su concreción.

RESUMEN

La vid (*Vitis vinifera*) es una planta frutal muy cultivada en el mundo debido a las considerables ganancias económicas asociadas a la industria vitivinícola. Por la riqueza de su tierra, clima y situación geográfica valiosa de cara a los puertos del pacífico, la Región de Cuyo es un lugar privilegiado para producir uvas y sus derivados de gran calidad. La producción argentina de uva se centra en las provincias de Mendoza y San Juan, y una gran parte se destina a la elaboración de vinos. Argentina se ubica como el octavo productor de uvas a nivel mundial según la Organización Internacional de la Viña y el Vino (OIV). La producción de jugo concentrado de uva (mosto) representa una fuente importante de ingresos económicos para las provincias de Cuyo, siendo exportado en su totalidad a Estados Unidos, Japón entre otros países. Por lo tanto, es evidente que para el fortalecimiento del comercio de uvas y productos derivados en los mercados internacionales más exigentes, deben asegurarse la autenticidad y los máximos estándares de calidad.

La calidad de los alimentos es un aspecto diferenciador en los mercados, ya que los consumidores ponen atención a determinados atributos de calidad de los productos y están dispuestos a pagar por ellos, dándoles cada vez más valor a la calidad y procedencia geográfica. Por ello, la Denominación de Origen Controlado (DOC) e Indicación Geográfica Protegida (IGP), se constituyen de manera creciente como parámetros aceptados internacionalmente para el reconocimiento de una calidad diferenciada, consecuencia de características propias y diferenciales, debidas al sitio geográfico de producción de alimentos, su elaboración e influencia del factor humano. Así, es necesario disponer de un sistema de "Trazabilidad", término que se define como el conjunto de acciones, medidas y procedimientos técnicos para identificar y registrar un producto desde su origen hasta la comercialización final. La trazabilidad otorga la posibilidad de comercializar productos en mercados más rentables, que exigen certeza en el origen y el proceso productivo. El INV ha desarrollado e implementado un sistema registral para la trazabilidad de vinos y uvas producidos en Argentina, que implica

documentar los procesos desde la cosecha hasta el destino final. Sin embargo, este sistema no contempla la identificación y/o determinación de parámetros químicos que permitan establecer la identidad inequívoca de las muestras en cualquier etapa de la cadena productiva, por lo que son vulnerables a maniobras fraudulentas de adulteración o contaminación durante su procesamiento.

Investigaciones recientes han demostrado que el análisis elemental es una herramienta muy valiosa para identificar el origen geográfico de vinos y clasificarlos de acuerdo a la variedad y origen de las uvas. Otros estudios han destacado la relación de los elementos químicos con el suelo, origen geológico y comportamiento de la vid. Dichas diferencias de concentración en los elementos químicos mayoritarios y traza se consideran como una "huella dactilar" de uvas y vinos que, desde el punto de vista quimiométrico, es útil para desarrollar modelos matemáticos conducentes a evaluar la trazabilidad. Las técnicas de "huella dactilar" involucran una variedad de métodos analíticos útiles para determinar la composición de productos alimentarios de manera no selectiva, esto es, mediante colección de espectros, cromatogramas o datos de composición multielemental.

Por todo lo anteriormente expuesto el objetivo general de esta tesis ha sido estudiar la presencia de posibles marcadores químicos de identidad en uvas y productos derivados de la Provincia de Mendoza y Argentina, con el fin de discriminar estos productos en función de su origen geográfico. Para ello, se evaluaron las concentraciones de diferentes elementos de la tabla periódica por ICP-MS en muestras de semillas y hollejos de uva y productos derivados como jugo de uva y vino. Las muestras de semillas y hollejos de uva y de vino se recolectaron de diferentes regiones vitivinícolas de importancia de Mendoza y otras regiones de Argentina, mientras que las muestras de jugos de uva abarcaron de países como Argentina y Brasil.

Todos los resultados fueron analizados mediante algoritmos estadísticos y de aprendizaje automático que permitieron modelar las variaciones presentes en los mismos y proponer modelos predictivos de origen geográfico para futuras muestras problema

desconocidas. En primera instancia se realizó un análisis exploratorio de datos mediante el uso de herramientas como análisis de componentes principales (PCA) o análisis de conglomerados (HCA), obteniéndose resultados que indicaron la necesidad de proponer modelos más complejos para agrupar las muestras de acuerdo a su origen geográfico. Por lo tanto, al considerar el sitio geográfico de obtención de las muestras, se hizo necesario utilizar técnicas de elevada complejidad, tales como máquinas de vectores soporte (SVM), el algoritmo de árboles aleatorios (RF), K vecinos más cercanos (K-NN), entre otros, para lograr resultados satisfactorios.

Los principales aportes de esta tesis se traducen en una profundización del conocimiento de la composición química mineral de semillas y hollejos de uva, como así también jugo de uva y vino de la Provincia de Mendoza y del resto de Argentina, abarcando no sólo la concentración de elementos de interés nutricional, sino también elementos a nivel de vestigios, logrando modelos quimiométricos que permiten establecer con mayor seguridad, la procedencia geográfica de muestras relevantes para el sector vitivinícola.

ABSTRACT

The vine (*Vitis vinifera*) is a worldwide cultivated plant due to the considerable economic gains associated with the wine industry. Owing to the richness of its land, climate and valuable geographical location, the Cuyo Region is a privileged place to produce high quality grapes and wines. Argentine grape production is centered in the provinces of Mendoza and San Juan, and a large part of it is destined to the production of wines. Argentina ranks as eighth producer of grapes worldwide according to the International Organization of Vine and Wine (OIV). The production of concentrated grape juice (must) represents an important source of economic income for the provinces of Cuyo, being exported entirely to the United States, Japan among other countries. Therefore, it is evident that in order to strengthen the trade of grapes and derived products in the most demanding international markets, authenticity and quality standards must be ensured.

The food quality is a differentiating aspect in the markets, since consumers pay attention to certain quality attributes of the products and are willing to pay for them, giving more value to quality and geographical origin. For this reason, the Denomination of Controlled Origin (DOC) and Protected Geographical Indication (PGI), are increasingly constituted as internationally accepted parameters for the recognition of a differentiated quality, as a consequence of their own and differential characteristics, due to the geographical site of food production, its preparation and influence of the human factor. Thus, it is necessary to have a "Traceability" system, a term that is defined as the set of actions, measures and technical procedures to identify and register a product from its origin to its final marketing. Traceability gives the possibility of commercializing products in more profitable markets, which require certainty in the origin and the production process. The National Institute of Viticulture (INV) has developed and implemented a registry system for the traceability of wines and grapes produced in Argentina, which involves documenting the processes from harvest to final destination. However, this system does not contemplate the identification or determination of

chemical parameters that make possible to establish the unequivocal identity of the samples at any stage of the production chain, therefore make them vulnerable to fraudulent adulteration or contamination maneuvers during processing.

Recent research has shown that elemental analysis is a very valuable tool to identify the geographical origin of wines and classify them according to the variety and origin of the grapes. Other studies have highlighted the relationship of chemical elements with the soil, geological origin and behavior of the vine. The differences in the major and trace chemical elements concentrations are considered as a "fingerprint" of grapes and wines which, from a chemometric point of view, is useful to develop mathematical models conducive to evaluating traceability. "Fingerprint" techniques involve a variety of analytical methods useful for determining the composition of food products in a non-selective manner, that is, by collecting spectra, chromatograms, or multi-element composition data.

For all the above, the main objective of this thesis has been to study the presence of possible chemical markers in grapes and products derived from the Province of Mendoza and Argentina, in order to discriminate these products based on their geographical origin. For this, the concentrations of different elements of the periodic table were evaluated by ICP-MS in samples of seeds and grape skins and derived products such as grape juice and wine. The samples of grape seeds and skins and wine were collected from different important wine regions of Mendoza and other regions of Argentina, while the samples of grape juices included from countries such as Argentina and Brazil.

Results were analyzed using statistical and machine learning algorithms that allowed modeling the variations present in them and proposing predictive models of geographic origin for future unknown problem samples. In the first instance, an exploratory data analysis was carried out using tools such as principal component analysis (PCA) or cluster analysis (HCA), obtaining results that indicated the need to propose more complex models to group the samples according to their geographic origin. Therefore, when considering the geographic

location for obtaining the samples, it became necessary to use highly complex techniques, such as support vector machine (SVM), random forest algorithm (RF), K nearest neighbors (K-NN), among others, to achieve satisfactory results.

The main contributions of this thesis are the knowledge of the mineral chemical composition of grape seeds and skins, as well as grape juice and wine from the Province of Mendoza and the rest of Argentina, covering not only the concentration of elements of nutritional interest, but also elements at the level of vestiges, achieving chemometric models that allow to establishing with greater security, the provenance of samples relevant to the wine sector.

PRODUCCIÓN CIENTÍFICA

La realización de la presente Tesis Doctoral ha dado origen a la redacción y publicación de diversos trabajos de investigación en revistas de prestigio internacional:

- ✓ **Intra-regional classification of grape seeds in Mendoza province (Argentina) by multi-elemental analysis and chemometric tools**

Brenda V. Canizo, Leticia B. Escudero, María B. Pérez, Roberto G. Pellerano y Rodolfo G.

Wuilloud

Food Chemistry 242 (2018) 272-278

Doi:10.1016/j.foodchem.2017.09.062

ISSN: 0308-8146

Editorial Elsevier

- ✓ **Data mining approach based on chemical composition of grape skin for quality evaluation and traceability prediction of grapes**

Brenda V. Canizo, Leticia B. Escudero, Roberto G. Pellerano y Rodolfo G. Wuilloud

Computer and Electronics in Agriculture 162 (2019) 514-522

Doi:10.1016/j.compag.2019.04.043

ISSN: 0168-1699

Editorial Elsevier

- ✓ **Provenance identification of commercial grape juices from Argentina and Brazil based on elemental composition and chemometrics techniques**

Brenda V. Canizo, Ana Laura Diedrichs, Agustín Londonio, Ezequiel Morzán, Patricia

Smichowski, Roberto G. Pellerano y Rodolfo G. Wuilloud

Enviado a: European Food Research and Technology

ISSN: 1438-2377

Editorial Springer

- ✓ **Geographical authentication of red wines from Argentina through multielemental analysis and chemometric tools**

Brenda V. Canizo, Lucila Brusa, Mirna Sigrist, Viviana P. Jofre, Roberto G. Pellerano y Rodolfo G. Wuilloud

En etapa de redacción.

Capítulo de libro:

- ✓ **Quality Monitoring and Authenticity Assessment of Wines: Analytical and Chemometric Methods.**

Brenda V. Canizo, Leticia B. Escudero, Roberto G. Pellerano, Rodolfo G. Wuilloud.

Libro: Quality Control in the Beverage Industry (Volume 17: The Science of Beverages). (2019) 335–384.

ISBN: 9780128166819

Doi:10.1016/B978-0-12-816681-9.00010-2.

Editorial Elsevier

Además, los trabajos resultantes se han presentado en diversos congresos nacionales e internacionales:

- ✓ **8º CONGRESO ARGENTINO DE QUÍMICA ANALÍTICA**; La Plata, Argentina. 2015.

“Análisis multielemental en uvas para vinificación cultivadas en la provincia de Mendoza por ICP-MS”.

Brenda Canizo; Leticia Escudero; Alexander Castro Grijalba; María Belén Pérez y Rodolfo Wuilloud.

- ✓ **14TH RIO SYMPOSIUM ON ATOMIC SPECTROMETRY**; Vitória, Espírito Santo, Brasil. 2017.

“Tracing the geographical origin of Mendoza (Argentina) grape seeds by ICP-MS multielemental analysis and advanced chemometrics techniques”.

Brenda Canizo; Leticia Escudero; María Belén Pérez; Roberto Pellerano y Rodolfo Wuilloud.

- ✓ **9º CONGRESO ARGENTINO DE QUÍMICA ANALÍTICA**; Río Cuarto, Argentina. 2017.

“Clasificación geográfica de uvas de Mendoza mediante análisis multielemental de hollejos y regresión logística multinomial”.

Brenda Canizo; Leticia Escudero; Alexander Castro Grijalba; Roberto Pellerano y Rodolfo Wuilloud.

- ✓ **10º CONGRESO ARGENTINO DE QUÍMICA ANALÍTICA**; La Pampa, Argentina. 2019.

“Caracterización de jugos de uvas comerciales en base al perfil mineral y análisis quimiométrico no supervisado”.

Brenda Canizo; Ana Laura Diedrichs; Agustín Londonio; Patricia Smichowski; Roberto Pellerano y Rodolfo Wuilloud.

- ✓ **15TH RIO SYMPOSIUM ON ATOMIC SPECTROMETRY**; Mendoza, Argentina. 2019.

“Classification of Mendoza (Argentina) red wines by ICP-MS multi-element analysis and multivariate statistic techniques”.

Brenda Canizo; Lucila Brusa; Mirna Sigrist; Roberto Pellerano y Rodolfo Wuilloud.

ÍNDICE GENERAL

CAPÍTULO I: INTRODUCCIÓN

1.1. Vitivinicultura	1
1.1.1. Historia de la vid y el vino	1
1.1.2. Argentina vitivinícola y desarrollo regional	4
1.1.3. Importancia comercial: datos estadísticos	10
1.1.4. Productos vitivinícolas	13
1.1.4.1. Uva	13
1.1.4.1.1. Características generales	13
1.1.4.1.2. Morfología y anatomía	14
1.1.4.1.3. Composición química	16
1.1.4.2. Jugo de uva	21
1.1.4.2.1. Características generales	21
1.1.4.2.2. Composición química	23
1.1.4.2.3. Proceso de elaboración	25
1.1.4.3. Vino	27
1.1.4.3.1. Características generales	27
1.1.4.3.2. Composición química	29
1.1.4.3.3. Proceso de elaboración	34
1.1.5. Referencias	37
1.2. Denominación de origen y trazabilidad	43
1.2.1. La importancia del origen de los alimentos	43
1.2.2. Definiciones y tipos de indicaciones geográficas: PDO, PGI y TSG	46
1.2.3. Enfoque vitivinícola: concepto de terruño o terroir	50
1.2.4. Reglamentación y marco legal	54
1.2.5. Referencias	56
1.3. Análisis multielemental para la designación de origen	59
1.3.1. El perfil elemental como 'fingerprint' de los alimentos	59
1.3.2. Técnicas analíticas elementales	65
1.3.2.1. Espectrometría de masas con plasma acoplado inductivamente	66
1.3.2.1.1. Instrumentación y funcionamiento	68
1.3.3. Referencias	73
1.4. Análisis quimiométrico para la designación de origen	76
1.4.1. Quimiometría e introducción al análisis multivariado	76
1.4.1.1. Datos multivariados	78
1.4.1.2. Pre-procesamiento de los datos	79
1.4.2. Análisis exploratorio de datos	82
1.4.2.1. Análisis de Componentes Principales (PCA)	83
1.4.2.2. Análisis Factorial (FA)	87
1.4.3. Análisis clasificatorio o de reconocimiento de patrones	87
1.4.3.1. Métodos no supervisados	90

1.4.3.2.	Métodos supervisados	94
1.4.3.3.	Validación del modelo	106
1.4.3.3.1.	Estrategias de validación	106
1.4.3.3.2.	Evaluación de los métodos quimiométricos	109
1.4.4.	Referencias	112

CAPÍTULO II: OBJETIVOS E HIPÓTESIS

2.1.	Objetivos generales	116
2.2.	Objetivos específicos	116
2.3.	Hipótesis del trabajo	116

CAPÍTULO III: REACTIVOS, MATERIALES Y MÉTODOS

3.1.	Reactivos, materiales e instrumentación	117
3.1.1.	Reactivos y materiales	117
3.1.1.1.	Agua ultrapura	117
3.1.1.2.	Soluciones patrón y estándares internos para mediciones en ICP-MS	117
3.1.1.3.	Reactivos generales	118
3.1.1.4.	Material de vidrio	118
3.1.2.	Instrumentación	118
3.1.2.1.	Espectrómetro de masas por plasma acoplado inductivamente	118
3.1.2.2.	Plancha calefactora	119
3.1.2.3.	Medición de microvolúmenes	119
3.2.	Muestreo y tratamiento de las muestras	119
3.2.1.	Hollejos y semillas de uva	119
3.2.2.	Jugos de uva	122
3.2.3.	Vinos	122
3.3.	Análisis multielemental mediante ICP-MS	124
3.3.1.	Hollejos y semillas de uva	124
3.3.2.	Jugos de uva	125
3.3.3.	Vinos	125
3.4.	Análisis estadístico quimiométrico	126
3.5.	Referencias	128

CAPÍTULO IV: RESULTADOS Y DISCUSIÓN

4.1.	Clasificación geográfica de uvas	129
4.1.1.	Clasificación intra-regional de semillas de uva de Mendoza mediante herramientas quimiométricas y análisis multi-elemental	129
4.1.1.1.	Introducción	129
4.1.1.2.	Determinación multielemental en semillas de uva	131

4.1.1.3.	Clasificación geográfica de semillas de uva de la provincia de Mendoza	137
4.1.1.3.1.	Análisis estadístico exploratorio	137
4.1.1.3.2.	Análisis estadístico clasificatorio	139
4.1.1.4.	Conclusiones	144
4.1.1.5.	Referencias	145
4.1.2.	Técnicas de minería de datos basado en la composición química de hollejos de uva para la evaluación de la calidad y la predicción de la trazabilidad de las uvas	149
4.1.2.1.	Introducción	149
4.1.2.2.	Concentración multielemental de hollejos de uva	151
4.1.2.3.	Análisis de PCA	155
4.1.2.4.	Análisis Predictivo	157
4.1.2.4.1.	Estimación de la exactitud de la clasificación	157
4.1.2.4.2.	Optimización de los algoritmos de clasificación	157
4.1.2.4.3.	Clasificación geográfica de hollejos de uva de la provincia de Mendoza	163
4.1.2.5.	Conclusiones	167
4.1.2.6.	Referencias	167
4.2.	Clasificación geográfica de jugos de uva	171
4.2.1.	Identificación de la procedencia de jugos de uva comerciales de Argentina y Brasil basados en la composición elemental y técnicas quimiométricas	171
4.2.1.1.	Introducción	171
4.2.1.2.	Concentración multielemental de los jugos de uva	173
4.2.1.3.	Análisis exploratorio	176
4.2.1.4.	Análisis clasificatorio	182
4.2.1.5.	Conclusiones	187
4.2.1.6.	Referencias	187
4.3.	Clasificación geográfica de vinos	191
4.3.1.	Autenticación geográfica de vinos tintos de Argentina mediante análisis multielemental y quimiométrico clasificatorio	191
4.3.1.1.	Introducción	191
4.3.1.2.	Concentración multielemental de vinos	194
4.3.1.3.	Análisis exploratorio	198
4.3.1.4.	Análisis clasificatorio	201
4.3.1.5.	Conclusiones	207
4.3.1.6.	Referencias	208

CAPÍTULO V: CONCLUSIONES Y PRESPECTIVAS FUTURAS

5.1.	Conclusiones y perspectivas futuras	212
-------------	--	------------

ÍNDICE DE TABLAS Y FIGURAS

TABLAS

Tabla 1.1. Posicionamiento de la Argentina en el contexto mundial vitivinícola	13
Tabla 1.2. Principales compuestos fenólicos de la uva	18
Tabla 1.3. Componentes químicos principales en uvas (% peso fresco)	20
Tabla 3.1. Detalles de las muestras de uvas analizadas	121
Tabla 3.2. Detalles de los vinos analizados	122
Tabla 3.3. Condiciones óptimas instrumentales de la determinación multielemental en semillas y hollejos	124
Tabla 3.4. Condiciones óptimas instrumentales de la determinación multielemental en jugos de uva	125
Tabla 3.5. Condiciones óptimas instrumentales de la determinación multielemental en las muestras de vinos	126
Tabla 4.1. Concentración multielemental de las muestras de semilla de uva según su origen geográfico (las concentraciones se expresan en $\mu\text{g/g}$ de base de materia seca)	132
Tabla 4.2. Resultados de la clasificación obtenidos con los diferentes modelos evaluados	143
Tabla 4.3. Concentración de los elementos determinados en las muestras de hollejo de uva según su origen geográfico (las concentraciones se expresan en $\mu\text{g/g}$ de base de materia seca)	152
Tabla 4.4. Comparación de la performance de los diferentes métodos de minería de datos aplicados en las muestras de prueba	164
Tabla 4.5. Resultados de la clasificación de los diferentes modelos aplicados a los conjuntos de entrenamiento y prueba	165
Tabla 4.6. Media y desviación estándar de la concentración de los elementos analizados en muestras de jugo de uva según su país de origen	173
Tabla 4.7. Concentración elemental de las muestras de vino según su región de procedencia	194

Tabla 4.8. Comparación de la performance de los diferentes métodos quimiométricos aplicados en las muestras de prueba	204
---	-----

FIGURAS

Figura 1.1. Cuadro al óleo representando a un joven dios Baco por Caravaggio, 1598	2
Figura 1.2. Provincias vitivinícolas de Argentina	5
Figura 1.3. Oasis vitivinícolas de la provincia de Mendoza	8
Figura 1.4. Destino de la producción de uvas	11
Figura 1.5. Estructura de una uva madura	16
Figura 1.6. Elaboración de jugo de uva	26
Figura 1.7. Principales pasos de la elaboración de vino tinto y blanco	36
Figura 1.8. Logos de DO e IG en el etiquetado de productos de Argentina	50
Figura 1.9. Enfoques para el monitoreo y evaluación de la designación de origen en vinos	63
Figura 1.10. Fuentes endógenas o primarias (línea continua) y exógenas o secundarias (línea punteada) de metales en el vino	64
Figura 1.11. Elementos que pueden determinarse por ICP-MS	67
Figura 1.12. Esquema de un instrumento ICP-MS	68
Figura 1.13. Diagrama esquemático de un ICP	70
Figura 1.14. Organización de los datos multivariados	78
Figura 1.15. Técnicas de pre-procesamiento de los datos	82
Figura 1.16. Esquema del proceso de reducción de variables en PCA	86
Figura 1.17. Enfoques de los métodos de reconocimiento de patrones	88
Figura 1.18. Proceso de clasificación	89
Figura 1.19. Representación ejemplo de un dendograma	94
Figura 1.20. Delimitadores para un caso bivariado simulado de LDA	96
Figura 1.21. Gráfico de un caso simulado de PLS-DA	98
Figura 1.22. Esquema de una red neuronal con una sola capa oculta	101
Figura 1.23. Esquema de los vectores de soporte	103
Figura 1.24. Descripción gráfica de cómo RF realiza la clasificación	106
Figura 1.25. Validación cruzada de k= 4 iteraciones	108
Figura 1.26. Ejemplo ilustrativo de una matriz de confusión	109
Figura 3.1. Ubicación geográfica de los viñedos de muestreo seleccionados	120

para la recolección de las muestras de uva	
Figura 4.1. Gráficos Box-plot de los elementos determinados en semillas de uva según su región de procedencia	136
Figura 4.2. Gráfico PCA de a) observaciones y b) variables de la PC1 vs. PC2 de las semillas	138
Figura 4.3. Gráfico de dispersión de las dos primeras funciones discriminantes del LDA de muestras de semilla de uva según su origen geográfico	140
Figura 4.4. Gráfico de caja y bigotes comparando los modelos aplicados para la clasificación de semillas de uva	144
Figura 4.5. Gráfico PCA de a) observaciones y b) variables de la PC1 vs. PC2 de los hollejos	156
Figura 4.6. Optimización de SVM	160
Figura 4.7. Número óptimos de árboles de clasificación para RF	161
Figura 4.8. Comparación de la optimización de mtry de RF por defecto (a) y grid (b)	162
Figura 4.9. Importancia de las variables en RF	166
Figura 4.10. Gráficos box-plot de los once elementos determinados en los jugos de uva de acuerdo al país de procedencia	175
Figura 4.11. Coeficientes de correlación de Pearson (r)	176
Figura 4.12. Gráficos de PCA de las observaciones (a) y variables (b) de la PC1 versus PC2. Importancia relativa de los elementos para las tres primeras componentes principales (c)	178
Figura 4.13. Dendograma del método de clúster jerárquico (a). Suma de los errores al cuadrado dentro del grupo (WSS) versus número de clústers (b). Gráfico de clúster (Clusplot) para el análisis de k-medias (c)	181
Figura 4.14. Rendimiento del modelo NNET al variar el tamaño de las unidades ocultas o neuronas y disminución de peso para los métodos de validación BTS (a) y CV (b)	184
Figura 4.15. Gráfico de las distribuciones de "accuracy" y "kappa" para los métodos de validación BTS (a) y CV (b)	186

Figura 4.16. Gráficos Box-plot de los elementos determinados en vinos tintos según su procedencia	196
Figura 4.17. Varianza explicada por las componentes principales	199
Figura 4.18. Gráfico PCA de a) observaciones y b) variables de la PC1 vs. PC2 de los vinos	200
Figura 4.19. Importancia de las variables en RF según la región geográfica	205
Figura 4.20. Gráficos box-plot comparando los modelos evaluados para la clasificación geográfica de vinos	206

ABREVIATURAS

AAS	Espectrometría de Absorción Atómica // Atomic Absorption Spectrometry
ADPIC	Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio
AM	Análisis Multivariado
AOC	Apelación de Origen Controlada // Appellation d'Origine Contrôlée
BTS	Método de remuestreo // Bootstrapping
CA	Corriente Alterna
CA	Análisis de Conglomerados // Cluster Analysis
CAA	Código Alimentario Argentino
CC	Corriente Continua
CV	Validación Cruzada // Cross Validation
DE	Desviación Estándar
DF	Funciones Discriminantes // Discriminant Function
DOC	Denominación de Origen Controlada // Denominazione di origine controllata
DRC	Celda de Reacción Dinámica // Dynamic Reaction Cell
EDA	Análisis Exploratorio de Datos // Exploratory Data Analysis
EM	Multiplicador de electrones // Electrons Multiplying
ETAAS	Espectroscopia de Absorción Atómica Electrotérmica // Electrothermal Atomic Absorption Spectrometry
FA	Análisis Factorial // Factorial Analysis
FAAS	Espectrometría de Absorción Atómica de Llama // Flame Atomic Absorption Spectrometry
FAES	Espectrometría de Emisión Atómica de Llama // Flame Atomic Emission Spectrometry
FN	Falsos Negativos
FP	Falsos Positivos
HCA	Análisis de Agrupamiento Jerárquico // Hierarchical Cluster Analysis
HF	Ácido Fluorhídrico
ICP-MS	Espectrometría de Masas por Plasma Acoplado Inductivamente // Inductively Coupled Plasma Mass Spectrometry

ICP-OES	Espectrometría de emisión óptica de plasma acoplado inductivamente // Inductively Coupled Plasma Optical Emission Spectrometry
IG	Indicación Geográfica // Geographical Indication
INAA	Análisis Instrumental de Activación de Neutrones // Instrumental Neutron Activation Analysis
INTA	Instituto Nacional de Tecnología Agropecuaria
INV	Instituto Nacional de Vitivinicultura
ISO	Organización Internacional de Normalización o Estandarización // International Organization for Standardization
K-NN	K-Vecinos más cercanos // K-Nearest Neighbors
LDA	Análisis Discriminante Lineal // Linear Discriminant Analysis
LOD	Límite de Detección // Limit of Detection
LR	Regresión Logística // Logistic Regression
m/z	Relación masa/carga
MLR	Regresión Logística Multinomial // Multinomial Logistic Regression
NNET	Redes Neuronales // Neural Networks
nt	Número de árboles / Number of trees
OIV	Organización Internacional de la Viña y el Vino
OMC	Organización Mundial del Comercio
OMPI	Organización Mundial de la Propiedad Intelectual
PC	Componente Principal // Principal Component
PCA	Análisis de Componentes Principales // Principal Component Analysis
PDO	Denominación de Origen Protegida // Protected Designation of Origin
PFA	Perfluoracetato // Perfluoroacetate
PGI	Indicación Geográfica Protegida // Protected Geographical Indication
PLS-DA	Análisis Discriminante de Mínimos Cuadrados Parciales // Partial Least Squares Discriminant Analysis
Potencia RF	Potencia de Radiofrecuencia
RBF	Función de Base Radial // Radial Basis Function
REE	Elementos de Tierras Raras // Rare Earth Elements
RF	Bosques Aleatorios // Random Forest

RFI	Importancia de Bosques Aleatorios // Random Forest Importance
RNAA	Análisis Radioquímico de Activación de Neutrones // Radiochemical Neutron Activation Analysis
SVM	Máquina de Vectores de Soporte // Support Vector Machine
TDS	Sólidos Disueltos Totales // Total Dissolved Solids
TNR	Tasa de Verdaderos Negativos // True Negative Rate
TPR	Tasa de Verdaderos Positivos // True Positive Rate
TSG	Garantía de Especialidad Tradicional // Traditional Speciality Guaranteed
UE	Unión Europea
VN	Verdaderos Negativos
VP	Verdaderos Positivos
WSS	Suma de los errores al cuadrado dentro del grupo // Within-cluster Sum of Squared error
XRF	Fluorescencia de Rayos X // X-Ray Fluorescence

CAPÍTULO I: INTRODUCCIÓN

1.1. VITIVINICULTURA

1.1.1. Historia de la vid y el vino

La viticultura, vinicultura o vitivinicultura (del latín *vitis*, "vid") es el arte y ciencia del cultivo de la vid, con el fin de utilizar las uvas como materia prima en la producción de vino u otros productos. Si bien su definición abarca la producción de otros productos derivados de la uva, su uso hace referencia principalmente al vino.

El conocimiento de la uva y el vino es tan antiguo como la historia de la humanidad. El vino tiene un registro arqueológico que data de más de 7500 años y los primeros vestigios de la presencia de vino datan desde principios hasta mediados del quinto milenio antes de Cristo, en la aldea Hajji Firuz Tepe, situada en la actual Irán [1,2]. La referencia escrita más antigua sobre el vino pertenece a la Biblia (Antiguo Testamento, Génesis, 9, 20-21) cuando Noé plantó vid después del diluvio, y luego hizo el vino [3]. Existen indicios de que el cultivo de la vid (al principio salvaje, denominada *Vitis vinifera sylvestris*) y la elaboración de bebidas a partir de las uvas ya se realizaban en torno a los años 6000 y 5000 años a.C., más específicamente fue en el Neolítico cuando se dieron las condiciones necesarias para su comienzo, en una vasta zona comprendida entre el mar Negro y el mar Caspio y delimitada por las actuales Turquía, Siria, Irak, Irán y Rusia. En el periodo Neolítico se desarrolló la primera gran revolución humana, ya que los hombres cambiaron definitivamente su estilo de vida nómada, después de miles de años de trashumar siguiendo las manadas de animales, y es cuando se desarrolla la agricultura, la metalurgia y se establecen las primeras ciudades [4]. Desde la Mesopotamia, el cultivo de la vid se extendió hacia el Cercano Oriente y hacia el Mediterráneo a través de las culturas babilónica, fenicia, griega, romana, etrusca y cartaginesa.

Desde Sumeria, la primera y más antigua civilización del mundo, el vino llegó a Egipto. Las orillas del Nilo fueron tierras de cultivo de la vid y en torno a estas se desarrolló toda una actividad laboral e industrial. Los egipcios fermentaban el mosto en grandes vasijas de barro y guardaban en ánforas, en las cuales los alfareros grababan quién había cultivado las uvas, la

fecha de elaboración y la calidad del mosto. El vino se convirtió en símbolo del estatus social y era empleado en ritos religiosos y festividades paganas [5].

El apogeo de la vitivinicultura antigua comenzó con los griegos. Se los considera los primeros “expertos en vinos” por los avances logrados en el cultivo y la vinificación. Fueron ellos los responsables de la difusión del consumo de vinos en toda la costa del Mediterráneo, hasta Portugal, el Norte de África y Asia Menor; también llevaron la vid a la actual Francia, habitada entonces por los galos [4]. Junto con la vitivinicultura, se desarrollaron en Grecia la alfarería y la orfebrería. Las ánforas de cerámica donde se fermentaba el vino eran semi-enterradas hasta que el vino estaba listo para la venta. Los griegos asignaron al vino una divinidad: Dionisio, que aparece siempre representado con una copa en la mano. En esta época, se elaboraban vinos con particularidades propias en diferentes regiones de Grecia, como en Rodas, Icaria, Quíos, Lesbos, Eritrea, Naxos, Taasos, Corinto o Mende.

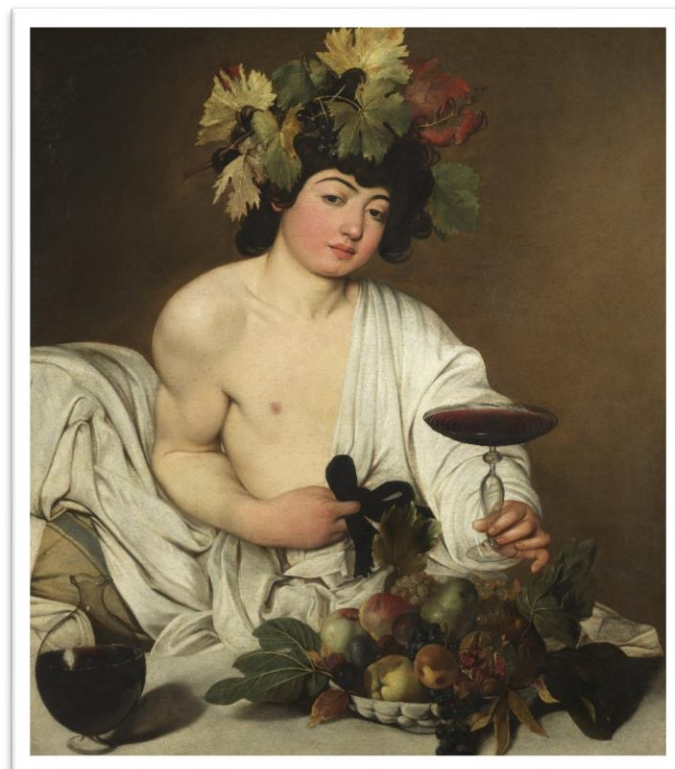


Figura 1.1. Cuadro al óleo representando a un joven dios Baco o Dionisio por Caravaggio, 1598.

Los romanos también fueron expertos viticultores y propagaron la actividad por todo el Imperio. La elaboración de vino se introduce en Italia en el 200 a.C. Los romanos adoptan al dios griego del vino cambiándole de nombre, así Dionisio se convierte en Baco, símbolo de la festividad asociada al consumo de vino. El primer mosto era mezclado con miel (obtenían lo que se conocía como *mulsum*, una apreciada bebida que se servía al comienzo de los banquetes), y el resto se almacenaba para que fermentara en grandes tinajas de barro. Aparece la figura del vinatero (antecesor del bodeguero moderno), quien añadía al vino sustancias para blanquearlos (los vinos blancos eran los más valorados por los romanos, de forma que se clarificaban), realizaba maceraciones con hierbas para proporcionarles aromas o guardaba parte de la cosecha en ánforas durante 15 o 25 años para que el vino madurase [4].

En Galia (Francia), se inventó la barrica de roble que reemplazó progresivamente a las ánforas romanas. La expansión del Imperio Romano llevó a que las tecnologías de viticultura y vinificación se extendieran hacia el oeste en la mayor parte de Europa, particularmente en España y Alemania, áreas ahora clasificadas como el "Viejo Mundo" del vino [6].

Con la caída del Imperio romano en el año 476 d.C., surge la Edad Media, y las tierras y con ellas los cultivos de vid, pasan a ser propiedad de la Iglesia y los reyes. Así, la elaboración del vino queda circunscrita a monasterios y castillos. En este periodo se hace extensivo el uso de las barricas de madera para almacenar el vino y, de forma casual, aparecen las primeras bodegas. Es gracias a la propagación del cristianismo que la viticultura debe su mayor desarrollo.

En la Península Ibérica, los Reyes Católicos iban reconquistando territorio a los musulmanes, terrenos en los que se replantaban vides. Entre los siglos XVII y XIX, los colonos españoles y británicos introdujeron tecnologías de viticultura y vinificación en países como Sudáfrica, Australia, Nueva Zelanda y América, países ahora clasificados como productores de vino del "Nuevo Mundo" [6]. En el continente americano ya existían variedades de *Vitis* silvestres (no viníferas). A estas variedades se unieron las traídas por los europeos.

Se abre una nueva etapa en la historia del vino, con el perfeccionamiento de su calidad en los siglos XVII y XVIII, y de las técnicas de vinificación. En este periodo, a medida que las ciudades crecían y aumentaba la riqueza de la burguesía, comenzó a crecer la demanda de vinos de más calidad, así los vinos de Borgoña, Burdeos y Champaña (Francia) adquieren parte de su posterior fama mundial, gracias, en gran medida, a los comerciantes del norte de Europa. Comienzan a utilizarse las botellas de vidrio para conservar el vino y se inventa el tapón de corcho. También en este momento, el monje Dom Pérignon descubre cómo elaborar el vino espumoso en la región de Champagne. La Iglesia católica fue responsable de la difusión de la uva y la inclusión cultural del vino en el nuevo mundo.

1.1.2. Argentina vitivinícola y desarrollo regional

La historia de la vitivinicultura Argentina se remonta a la época de la colonización, ya que el cultivo de la vid estaba estrechamente relacionado con las prácticas agrícolas de los colonos españoles que trajeron *Vitis vinífera* con la finalidad de obtener vinos para celebrar las misas religiosas. Las primeras especies de *Vitis vinífera* llegaron a mediados del siglo XVI al Cuzco (Perú), de allí pasaron a Chile y, a partir de 1551 fueron introducidas en la Argentina, y gracias a las fundaciones de Santiago del Estero (1553) y de Mendoza (1561) que abrieron el camino para la introducción de las primeras cepas de vid, que se propagaron por Cuyo, Tucumán y el área Rio Platense hasta consolidarse en los siglos XVII y XVIII. Fue la etapa de la vitivinicultura criolla, con centro principal en Mendoza y mercados de consumo en Córdoba, Santa Fe y Buenos Aires, entre otros [7].



Figura 1.2. Provincias vitivinícolas de Argentina [4].

La zona de la Argentina en donde se desarrolló la vitivinicultura fue principalmente en la zona pedemontana de la Cordillera de Los Andes, en regiones que poseían casi exclusivamente un clima árido desértico. En la Figura 1.2 se observan las principales provincias en las que hasta la actualidad sigue en actividad la vitivinicultura. El cultivo de la vid se instaló en regiones climáticas aptas a lo largo de la Cordillera de los Andes, entre los 22° y los 42° de latitud Sur y con altitudes entre 500 m y 1500 m sobre el nivel del mar. La gran amplitud latitudinal, combinada con la topografía de los valles andinos, generaron grandes variaciones ecológicas que permitieron el cultivo de una amplia gama de cepas, dando diferente potencial de producción vitícola a cada zona o región. La distancia de la nueva ciudad respecto de Santiago de Chile y la necesidad del vino para consumo diario y para las misas, incentivó la vitivinicultura. A fines del siglo XVI, ésta dejó de ser artesanal y doméstica para transformarse en una actividad económica de importancia.

Hacia el año 1600, Mendoza elaboraba una importante cantidad de vinos cuyos primeros destinos fueron a nivel nacional y tiempo más tarde llegaron las primeras exportaciones hacia Chile, España e Inglaterra [8]. Con el correr de los siglos y gracias a las inmigraciones europeas que se instalaron en la provincia de Mendoza con conocimientos del cultivo de la vid y de la elaboración de vinos, se comenzó a cultivar diversas variedades de vid para la producción de vinos finos, logrando la especialización vitivinícola de Cuyo. En 1853, con el apoyo del presidente Domingo Sarmiento, se fundó la Quinta Normal para la Enseñanza de la Agricultura y Técnica en Mendoza, para la cual, el gobierno de la provincia contrató al francés Michel Pouget para dirigir la institución. Pouget trajo de Francia centenares de nuevas especies forestales, además de métodos científicos para su cultivo, como poda racional e injertos, y maquinarias e implementos agrícolas. La vitivinicultura continuó un ininterrumpido camino de perfeccionamiento y calidad, lo que llevaría a convertir a Mendoza en una de las “grandes capitales mundiales del vino”.

Otras regiones desarrollaron la vitivinicultura permitiendo que diferentes variedades de vid pudieran encontrar mejores condiciones para su expresión, creando un rico abanico de opciones para la vitivinicultura argentina. En la provincia de Salta y Catamarca, la viticultura se desarrolló en la región de los Valles Calchaquíes, el cual se extiende en estas dos provincias, con el varietal blanco Torrontés Riojano como la variedad más extendida. Igualmente, entre los tintos se encontraron el Malbec, Syrah, Merlot y el Cabernet Sauvignon [4]. Esta región concentra actualmente más de 4.000 hectáreas de viñedos [9] y se caracteriza por ser una región de clima seco, y por una gran amplitud térmica con veranos largos que permiten un buen crecimiento de la vides favorecido también por suelos arenosos y profundos [10]. En la provincia de La Rioja el desarrollo comenzó en las regiones formadas por dos formaciones de cerros, las cadenas de Famatina y Velasco, que dan lugar a dos valles importantes: el Valle de Famatina, entre las dos cadenas, y el del Bermejo, que se extiende entre el Famatina y la Pre-Cordillera de Los Andes. En estas regiones existen más de 7500 hectáreas de viñedos,

concentradas fundamentalmente en la localidad de Chilecito, y es el Torrontés Riojano el cepaje blanco predominante. Entre los tintos, las variedades Malbec, Syrah y Bonarda lograron una interesante adaptación a la zona. San Juan es la segunda provincia en importancia luego de Mendoza, con casi 46.000 hectáreas de viñedos. Esta región está formada por cuatro valles importantes: el Valle del Tulum, Valle del Pedernal, Valle del Ullum y Valle del Zonda, en las cuales son típicas las variedades blancas como Torrontés Sanjuanino, Moscatel de Alejandría y Pedro Ximénez. También se destacan las uvas rosadas de la variedad Cereza y entre los cepajes tintos se ha revalorizado Bonarda y el cepaje Syrah. Otras variedades blancas como Chardonnay, Torrontés Riojano, Sauvignon Blanc y Viognier, también se destacan en la región. Además de la producción de vino, San Juan es un importante productor de jugo de uva concentrado y es la principal provincia en la producción y exportación de uvas destinadas al consumo en fresco y pasas [4]. Finalmente, las provincias de Neuquén y Río Negro, constituyen la zona vitivinícola más austral, ubicada en la Patagonia Argentina. El cultivo de la vid se concentra en el Alto Valle del Río Negro, formado por los ríos Negro, Limay, Neuquén y Colorado. Ambas provincias cuentan con más de 1500 hectáreas de superficie cultivada [9]. Por sus características agroclimáticas, Neuquén y Río Negro presentan las condiciones óptimas para el desarrollo de variedades de zonas frías como las tintas Merlot y Pinot Noir, aunque también se pueden encontrar varietales como Malbec. Entre los blancos se destacan Semillón y Sauvignon Blanc. Las variedades como Traminer y Riesling también caracterizaron durante mucho tiempo a la región.

En general, las regiones vitivinícolas argentinas se caracterizan por tener pocas lluvias y baja humedad atmosférica con inviernos bien marcados, veranos calurosos y buena insolación, lo que permite una completa maduración de las uvas y por ende una buena tipicidad varietal.

Mendoza es el centro vitivinícola más importante del país, concentra casi el 70% de la producción de uvas y vinos de Argentina con más de 150.000 hectáreas, y la mayor cantidad de bodegas del país. Los suelos pobres, una notable amplitud térmica, escasas lluvias, la buena

insolación de las uvas y la particularidad de desarrollar un “oasis” de riego que recibe agua de deshielo de los glaciares andinos cercanos, son factores que caracterizan esta región vitivinícola. Se distinguen en Mendoza cinco zonas bien diferenciadas que corresponden a los distintos oasis. Entre ellos, encontramos la *Zona alta del Río Mendoza*, la *Región Norte*, luego encontramos la *Región Este*, el *Valle de Uco*, y finalmente la *Región Sur* de Mendoza [4,10].

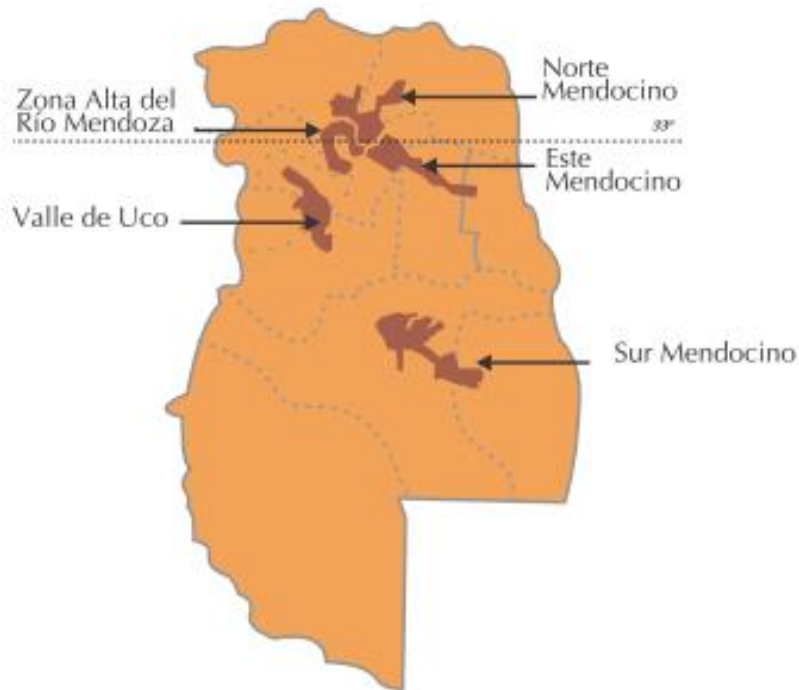


Figura 1.3. Oasis vitivinícolas de la provincia de Mendoza [4].

La *Zona alta del Río Mendoza* corresponde a una zona vitícola que está ubicada en el piedemonte y regada por el Río Mendoza. Abarca áreas vitícolas de los Departamentos Luján de Cuyo, Maipú, Guaymallén y Las Heras. La temperatura media anual que es de unos 15°C y la marcada amplitud térmica son factores que otorgan personalidad a las variedades mejor adaptadas a la región. El suelo se caracteriza por la presencia de cantos rodados a diferentes profundidades. Se caracteriza por la elaboración de vinos de calidad.

La *Región Norte* es una planicie que se extiende en los costados del bajo Río Mendoza y comprende las áreas de menor altura sobre el nivel del mar. Abarca el departamento de Lavalle y parte de los departamentos de Maipú, Guaymallén, Las Heras y San Martín. En

general, son terrenos de poca pendiente y profundos. El Syrah y el Bonarda son cepajes característicos de la región. También, la región se caracteriza por la producción de vinos blancos, y entre las variedades más extendidas se encuentran el Chenin Blanc, Pedro Ximénez, Ugni Blanc y Torrontés.

La *Región Este* está conformada por los departamentos San Martín, Junín, Santa Rosa, La Paz y Rivadavia, y representa la mayor región productora de vinos del país. Es una planicie que recibe las aguas del Río Tunuyán y parte del Río Mendoza. Tiene un clima templado cálido y cuenta con vastas horas de sol favorables para el desarrollo vegetativo, lo cual permite una óptima maduración de las variedades cultivadas. Tiene una marcada amplitud térmica durante todo el año. Sus suelos son de origen eólico, de textura franco-arenosa y areno-limosa, ricos en potasio. En los últimos años, la región ha mostrado un notable progreso en el manejo de los cultivos, sistematización del riego, incorporación de tecnología en viñedo, bodega y en establecimientos dedicados a la elaboración de jugo de uva concentrado. Las principales variedades para vinos tintos son Bonarda, Cabernet Sauvignon, Sangiovese, Merlot y Tempranillo y para vinos blancos: Pedro Jiménez y Ugni Blanc.

La *Región del Valle de Uco*, con gran desarrollo vitivinícola en la actualidad, está conformada por los departamentos Tunuyán, Tupungato y San Carlos, ubicados en el Centro-Oeste de Mendoza. Su clima es templado con inviernos rigurosos y veranos cálidos, de noches frescas. Los cepajes blancos de mayor desarrollo en la región son: Chardonnay, Semillón, Torrontés y Pedro Ximenez. Entre los tintos se destacan Malbec, Merlot, Cabernet Sauvignon, Tempranillo y Bonarda. Y en las zonas más altas variedades como Pinot Noir y Syrah, encuentran las condiciones adecuadas para su óptima maduración. La región se caracteriza por su aptitud para la producción de materia prima de excelente calidad, tanto para la obtención de vinos blancos como tintos.

Finalmente, la *Región Sur* comprende los departamentos de San Rafael y General Alvear, y se extiende al pie de la cordillera principal. Es un oasis irrigado por las aguas de los

ríos Atuel y Diamante [11]. Esta área ofrece excelentes condiciones climáticas y edáficas para el cultivo de la vid. Los suelos son aluvionales presentando gran proporción de calcáreo. San Rafael es el segundo departamento de la provincia en importancia de producción vitivinícola. Las principales variedades para vinos tintos son Bonarda, Cabernet Sauvignon, Merlot, Syrah y Malbec y para vinos blancos: Chenin Blanc, Semillón, Torrontés y Chardonnay, utilizados para base de espumantes.

1.1.3. Importancia comercial: datos estadísticos

Anualmente el panorama mundial sobre la situación vitivinícola se expone en un informe elaborado por la Organización Internacional de la Viña y el Vino (OIV), en el mismo, se estima que en el año 2018 el área mundial bajo viñedos, que corresponde a la superficie total plantada con viñedos, incluida la que aún no está en producción y para todos los fines (vino, mesa y pasas), es casi equivalente al área del año 2017, llegando a 7,4 millones de hectáreas, un número que refleja la importancia que tiene la vitivinicultura a nivel mundial. Del total de la superficie destinada al cultivo de la vid 5 países representan el 50% de la viña mundial, España (13%), China (12%), Francia (11%), Italia (9%) y Turquía (6%) [12]; Argentina ocupa el séptimo puesto con una superficie destinada a la vid de 218.000 hectáreas. En el país hay 18 provincias que registran superficie cultivada de vid, siendo principalmente 7 las que concentran la mayor superficie. El 70,1% del total se encuentra en Mendoza, 21,4% en San Juan, 3,6% en La Rioja, 1,5% en Salta, 1,3% en Catamarca, 0,8% Neuquén y 0,7% Río Negro. El restante 0,5% se distribuye en 11 provincias: Córdoba, La Pampa, Buenos Aires, Tucumán, San Luis, Chubut, Entre Ríos, Jujuy, Misiones, Santiago del Estero y Santa Fe [9].

Con respecto a la producción mundial de uvas, en el año 2018 fue de 77,8 millones de toneladas, de las cuales el 57% se destinó a las elaboraciones de vino, el 36% a uvas de mesa y el 7% a la desecación para pasas [12]. Argentina se ubicó como octavo productor de uvas con una cosecha que alcanzó casi los 26 millones de quintales, manifestando un crecimiento del 31% respecto al año anterior, de ese total aproximadamente 17 millones de quintales de uvas

fueron cosechados en la provincia de Mendoza, correspondiendo un 67% [13]. Considerando el destino de las uvas ingresadas a establecimientos de la producción nacional graficada en la Figura 1.4, se observa que se destinó a bodegas y/o fábricas de mosto un 97,8% de uvas, a secaderos de pasas un 2% y a productores de uva para consumo en fresco sólo un 0,2% [13].

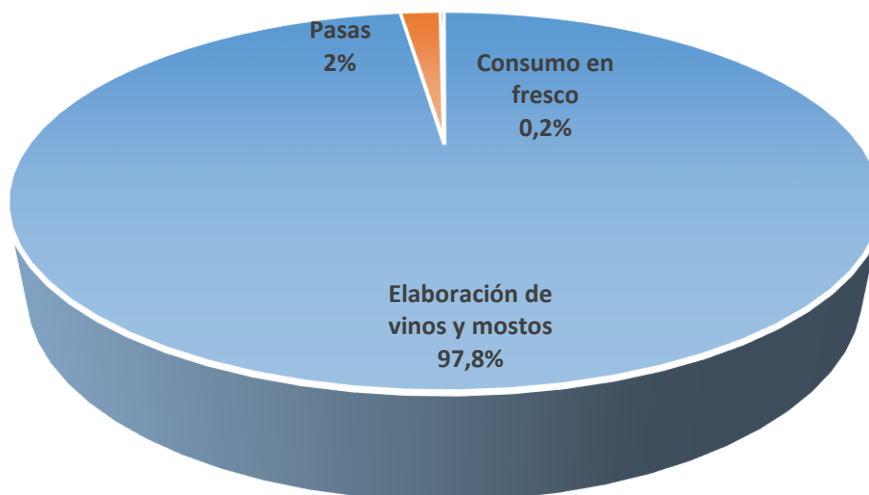


Figura 1.4. Destino de la producción de uvas de Argentina [13].

La tendencia de producción de uva para consumo en fresco es decreciente desde hace varios años. Con respecto a la uva destinada a pasas en 2018 se observó un incremento de un 34% más que en el año anterior, ubicando a Argentina como octavo productor mundial de uvas pasas.

En cuanto a la producción global de vino, el principal producto de destino de la uva, según el informe anual de la OIV la elaboración de vino (excluyendo jugos y mostos) se estimó en 292,3 millones de hectolitros en el 2018, un aumento marcado de 42,5 millones de hectolitros en comparación con la producción de 2017 [12]. Argentina se ubica como quinto productor mundial de vino, en el año 2018 se elaboraron 14,5 millones de hectólitros de vino, casi 5 millones de hectólitros de mosto y 1631 hectólitros de jugo de uva. Estos valores indican un aumento del 23% en la elaboración de vinos y 73% en la elaboración de mostos y jugo de uva respecto al año 2017. La provincia de Mendoza es la principal productora de vino del país, en el año 2018 elaboró casi 11 millones de hectolitros, pero con respecto a la elaboración de

mostos de uva, San Juan es el primer productor del país, elaborando para el año 2018, un volumen de 2 millones de hectolitros [13]. El jugo concentrado de uva o mosto es un producto estratégico para la industria vitivinícola porque toma, aproximadamente, el 25% del volumen total de la cosecha, lo que permite “equilibrar” los stocks de vinos genéricos e influir positivamente sobre sus precios. De la producción total de mosto en Argentina, el 90% del volumen es exportado principalmente a cuatro destinos, Estados Unidos, Sudáfrica, Japón y Canadá, que representan más del 80% del total exportado.

El consumo mundial de vino en 2018 se estima en 246 millones de hectolitros, en este sentido Argentina no es sólo un importante país productor, también es un gran consumidor de vinos, por este atributo se ubica en el noveno lugar a nivel mundial. Respecto a la exportación de vino, Argentina se ubicó en el 2018 como décimo exportador a nivel mundial, observándose un notable aumento de más del 23% con respecto al año anterior. Si se considera sólo América Latina, Argentina ocupa el segundo lugar detrás de Chile, quien tiene un perfil marcadamente exportador. Las exportaciones de vino fraccionado de Argentina en 2018 tuvieron como principal destino Estados Unidos, con un 27% del volumen exportado, el Reino Unido aparece como segundo destino con 13%, seguido por Brasil, Canadá y Paraguay con 9%, 8% y 6% respectivamente. Con respecto a las exportaciones de mosto, el principal destino fue Estados Unidos con el 33% del total, lo sigue Sudáfrica con 20% y Japón con 19%.

A modo de resumir la posición de Argentina en el contexto mundial de la vitivinicultura, se presenta la Tabla 1.1; en ella se puede apreciar la importancia de nuestro país dentro de esta industria, ubicándose entre las diez principales a nivel mundial.

Tabla 1.1. Posicionamiento de la Argentina en el contexto mundial vitivinícola.

Detalle	Posición
Superficie	7
Producción de uva	8
Elaboración de vinos	5
Consumo de vinos	9
Exportaciones de vinos	10

1.1.4. Productos vitivinícolas

1.1.4.1. Uva

1.1.4.1.1. Características generales

La vid (*Vitis vinifera*) es uno de los cultivos de mayor importancia para la economía mundial. La mayor parte de la producción de su fruta, la uva, se destina a la elaboración de vino, aunque también porciones significativas se consumen frescas, secas en pasas, procesadas como jugo sin alcohol y destiladas en bebidas espirituosas [14].

Botánicamente, la vid es una planta semileñosa y/o trepadora que se clasifica en la familia *Vitaceae*, en el género *Vitis*, que a su vez posee dos subgéneros, *Muscadinia* y *Euvitis*; es en éste último subgénero al cual pertenece la especie *Vitis vinifera* y su antecesora *Vitis silvestris*. *Vitis vinifera* tiene infinidad de variedades que no sólo tienen interés botánico, sino también agrícola, las cuales se nombran como variedad, cepaje o cultivar. En esta especie se encuentran todas las variedades utilizadas para la elaboración de vino, jugo y otros usos industriales, además de las variedades para consumo en fresco.

La vid se desarrolla adecuadamente en regiones de clima templado, con veranos largos, cálidos y secos e inviernos fríos. Es una especie criófila, las temperaturas bajo cero son muy convenientes para su reposo vegetativo, aunque las heladas tardías provocan daños sobre las

estructuras en crecimiento. La producción de uva está restringida en gran medida a regiones climáticas caracterizadas por climas de tipo mediterráneo, se aproxima al área entre las isothermas anuales de 10 y 20°C [1].

La radiación solar es importante para la acumulación de azúcares en el fruto, sin embargo esa radiación solar solo es eficaz si es interceptada por el follaje. Esto depende del sistema de conducción. En cuanto a la humedad, la vid requiere bajos niveles de humedad ambiental para evitar la proliferación de enfermedades criptogámicas que afectan su desarrollo y la calidad de los frutos. Se puede realizar el cultivo en distintos tipos de suelos, ya que se adapta a diversas condiciones edáficas, pero los óptimos para su crecimiento son aquellos de textura franca, moderadamente profundos a profundos, de buen drenaje y adecuada fertilidad.

1.1.4.1.2. Morfología y anatomía

Botánicamente, la uva es una baya carnosa y jugosa que está unida a un racimo. Cuando hablamos de la morfología de la planta de la vid se distinguen claramente dos partes. La subterránea, formada por las raíces de mayor o menor grosor, cuyas extremidades más finas y jóvenes forman la cabellera. En la parte aérea se distinguen el tronco, los brazos y los sarmientos por una parte y los brotes, hojas, frutos y zarcillos por otra. Esta parte aérea es lo que la mayoría de los autores denominan canopia [4].

El crecimiento de la baya se inicia luego de una exitosa polinización y fertilización de los óvulos dentro de la flor. La formación y el crecimiento de las bayas de uva (*Vitis vinifera*) siguen un patrón sigmoide doble con tres fases distintas: I, división celular rápida y expansión en bayas verdes; II o fase de retraso, en la que cesa la expansión celular; y III, en el que se reinicia el crecimiento y el fruto madura [15,16]. Desde la floración hasta aproximadamente 60 días después, se produce la primera fase de crecimiento rápido durante la cual se forma la baya y se producen los embriones de semillas. A medida que se desarrollan las uvas, muestran modificaciones en tamaño, composición, color, textura, sabor y susceptibilidad a los patógenos.

En el periodo I en el que la baya forma su estructura y desarrolla las semillas, se produce y almacena una gran cantidad de ácidos orgánicos en la vacuola de las células de pericarpio. El cambio en el color de la piel, que indica el comienzo de la maduración, se ha adoptado para describir el inicio de la maduración. Los cambios más dramáticos en la composición de las uvas se producen durante esta segunda fase de crecimiento o fase de maduración. Las bayas cambian de un estado en el que son pequeñas, duras y ácidas, con poco azúcar a un estado en el que son más grandes, más suaves, más dulces, menos ácidas y con un fuerte sabor y color. Además existe acumulación no sólo de grandes cantidades de azúcares, sino también de productos secundarios como compuestos aromáticos y polifenoles como las antocianinas [17].

En la Figura 1.5 se presenta la estructura de una baya de uva madura. En la estructura de la uva se pueden distinguir dos partes claramente diferenciadas, las semillas y el pericarpio o conjunto de tres tipos de tejidos organizados concéntricamente alrededor de las semillas. Se puede observar que el grano de uva está formado por el hollejo o piel que botánicamente corresponde al epicarpio o exocarpio y que está recubierto de una capa cerosa llamada pruina o cutícula. El exocarpio, consiste en una epidermis cubierta por cutículas, y varias capas de células subyacentes de hipodermis de paredes gruesas, contiene la mayoría de los flavonoides y antocianinas en las capas más externas de las variedades de uva roja. La cutícula de la baya de uva madura tiene una región externa amorfa y una región interna principalmente reticulada. Hacia adentro sigue la pulpa (mesocarpio) consiste en parénquima de pared delgada. Las células son redondas a ovoides y contienen grandes vacuolas, que son los sitios principales para la acumulación de azúcares, compuestos fenólicos, agua y ácidos orgánicos durante la maduración de la uva [15]. Por último encontramos las semillas o pepitas (endocarpio) que son los óvulos fecundados y que como máximo se encuentran cuatro. Cada una de estas partes tiene su estructura celular y propiedades bioquímicas. El pincel, formado por la continuación de los vasos del pedicelo, permite a la baya nutrirse. En las células del hollejo se encuentra la

materia colorante, que en las variedades tintas también aparece en la pulpa. Los perfumes varietales o primarios, se encuentran en las primeras capas subyacentes (hipodermis) y mucho menos en las primeras capas de células de la pulpa [15–17].

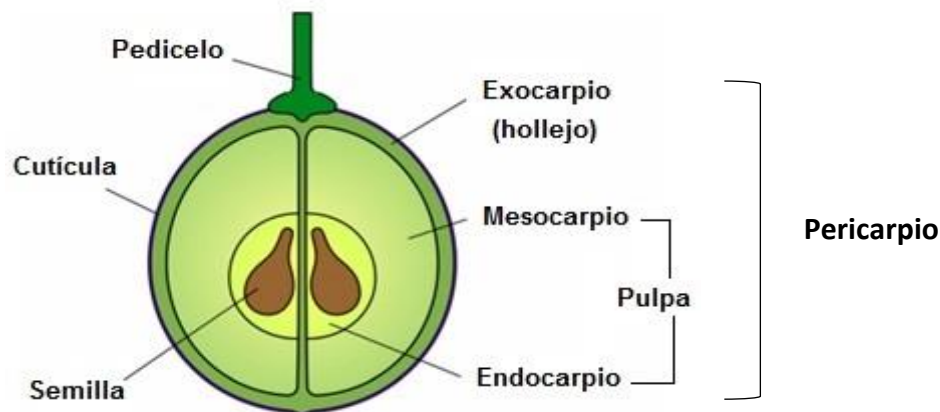


Figura 1.5. Estructura de una uva madura [16].

Los racimos están formados por el escobajo, raquis o raspón y los granos o bayas. En líneas generales, los porcentajes aproximados (dados en peso) de cada una de las partes son: escobajo 5%, hollejo 7%, pulpa 84% y semillas 4% [7].

1.1.4.1.3. Composición química

La composición química de la uva es compleja y contiene cientos de compuestos. El agua (75–85%) es el componente principal seguido de azúcares y luego ácidos orgánicos. Otros compuestos importantes incluyen aminoácidos, proteínas y compuestos fenólicos [15].

Carbohidratos

Los principales azúcares de la uva son las hexosas, glucosa y fructosa y en cantidades relativamente insignificantes otros monosacáridos (pentosas). El disacárido, sacarosa, rara vez se encuentra en las uvas *Vitis vinifera*, pero puede constituir hasta el 10% del contenido de azúcar en los cultivares no vinifera. A menudo se encuentran en proporciones aproximadamente iguales en la madurez, mientras que las uvas demasiado maduras tienen una mayor proporción de fructosa. El contenido de azúcar es un indicador importante que se

utiliza para evaluar la madurez e indicar el momento óptimo de cosecha de la uva. El contenido de azúcar en la uva varía dependiendo de la especie, variedad, madurez y salud de la fruta. Los cultivares de *V. vinifera* generalmente alcanzan una concentración de azúcar del 20% o más por madurez [18]. Otros azúcares como rafinosa, estaquiosa, melibiosa, maltosa, galactosa, arabinosa y xilosa también se acumulan en cantidades insignificantes en las uvas maduras.

Otros compuestos encontrados en las uvas son las pectinas, que son heteropolisacáridos que funcionan como el pegamento primario que mantiene juntas las células de la uva. Su retención al final de la maduración proporciona a las uvas de mesa su textura tradicional. En contraste, la hidrólisis de pectina es crítica para el ablandamiento de las uvas de vino. El ablandamiento facilita la liberación de jugo y sabor durante la trituración de la uva y minimiza la probabilidad de producción de turbidez en los vinos.

Ácidos orgánicos

Los ácidos orgánicos tartárico y málico constituyen alrededor del 90% del contenido total de ácidos orgánicos en las uvas maduras. Cantidades menores de ácido cítrico, succínico, láctico y acético también están presentes en las uvas. El resto de los ácidos consiste en cantidades variables de ácidos fenólicos (por ejemplo, ácidos quínico y shikímico), aminoácidos y ácidos grasos.

Compuestos fenólicos

Los compuestos fenólicos constituyen el tercer grupo más significativo de compuestos orgánicos que se encuentran en las uvas. Estos compuestos son muy importantes para la calidad del vino porque son responsables de la mayoría de las características sensoriales, particularmente el color y la astringencia. Estos grupos de compuestos constituyen un grupo diverso de metabolitos secundarios que existen en las uvas, principalmente en las pieles y semillas de las uvas y también en los tallos de las uvas. Los compuestos fenólicos en las uvas de *V. vinifera* incluyen dos clases de compuestos fenólicos: flavonoides y los no flavonoides. Los

compuestos no flavonoides incluyen ácidos fenólicos y también otros derivados fenólicos como los estilbenos. Para los compuestos flavonoides, hay una gran cantidad de subclases, como flavonoles, flavanoles y antocianinas. En la piel u hollejo, los isómeros de resveratrol (cis y trans) junto con las antocianinas (pigmentos de colores) son los compuestos fenólicos más abundantes, mientras que en las semillas predominan los flavanoles, a saber (+)-catequina y (-)-epicatequina, mientras que la pulpa es rica en ácidos fenólicos. En general, la composición fenólica de las uvas está influenciada por diferentes factores, como la variedad de uva, la exposición a la luz solar, la radiación solar, la altitud, la composición del suelo, el clima, las prácticas de cultivo, exposición a enfermedades y el grado de madurez de la uva [15]. La Tabla 1.2 resume los principales compuestos fenólicos de la uva.

Tabla 1.2. Principales compuestos fenólicos de la uva.

Compuestos Fenólicos		Localización
No flavonoides	Ácidos fenólicos	Ácidos hidroxicinámicos presentes como ésteres tartáricos Ácidos benzoicos (principalmente ácido gálico)
	Estilbenos	Resveratrol
Flavonoides	Antocianinas	Antocianidinas: cianidina, peonidina, delfinidina, petunidina, malvidina, (en forma de glucolizado y eventualmente acilado derivados)
	Flavan-3-oles	Monómeros(catequinas, galocatequina, epicatequina, epigalocatequina)
		Oligómeros y polímeros o proantocianidinas (procianidinas, prodelfinidinas)
	Flavonoles	Campherol, quercetol, miricetol, isorhamnetol
Flavanonoles	Engeletin, astilbin	

Compuestos nitrogenados

Los compuestos nitrogenados de las uvas incluyen cationes de amonio y compuestos orgánicos nitrogenados como aminoácidos, hexosa aminas, péptidos, ácidos nucleicos y

proteínas. A medida que ocurre la maduración, el nitrógeno orgánico aumenta progresivamente mientras que el amoníaco disminuye ligeramente. La concentración de estos compuestos es relativamente baja y tiene poco efecto directo sobre el sabor o el flavor. En las uvas, los principales aminoácidos libres incluyen la prolina, la arginina y, en menor medida, la alanina, ácido aspártico y ácido glutámico. La variación del perfil de aminoácidos y su concentración en las uvas depende de la variedad de la uva, pero también del manejo vitícola y las condiciones ambientales. [15,16,18].

Compuestos del aroma

Con respecto al aroma y los componentes del flavor, los compuestos que otorgan a variedades particulares de uva su aroma distintivo generalmente se presentan en pequeñas cantidades, ocasionalmente existiendo en pocas partes por billón. Por lo general, se acumulan a medida que las uvas maduran. A medida que se acumulan, una fracción creciente puede unirse en complejos no volátiles, como glucósidos, polímeros o derivados oxidados. Esto es particularmente notable con aromatizantes como monoterpenos, norisoprenoides y fenólicos no flavonoides [15,18].

Compuestos lipídicos

El componente lipídico de las uvas consiste principalmente en ceras superficiales (principalmente ácido oleanólico) y ácidos grasos cutáneos, fosfo- y glicolípidos de membrana y aceites de semillas. De estos, los más significativos son los ácidos grasos de cadena larga en la superficie de la piel. Estos con frecuencia actúan como fuentes importantes de ácidos grasos insaturados durante la fermentación del vino. Aunque se producen en pequeñas cantidades, los derivados de carotenoides oxidados de las uvas generan importantes compuestos aromáticos.

Minerales

Finalmente la uva contiene numerosas sustancias minerales que la planta extrae del suelo y son transportados por el agua con otros elementos necesarios para su crecimiento. La

concentración iónica de esta solución es función de la naturaleza del suelo del viñedo y la fertilización [19]. Las sustancias minerales se localizan sobre todo en las partes sólidas de la uva: hollejo, semillas, paredes celulopecticas de las células de la pulpa, mientras que el jugo vacuolar, es decir el mosto, es proporcionalmente menos rico. A lo largo de su crecimiento (fases herbáceas y/o de maduración) la baya acumula grandes cantidades de nitrógeno, potasio, calcio, fósforo y magnesio. Las uvas son muy ricas en potasio, que es un macronutriente esencial para el crecimiento y desarrollo de la vid y la baya de uva. En la uva madura, la mitad del potasio se localiza en la piel. Otros elementos, como el calcio y el magnesio, junto con cantidades mínimas de sodio y hierro, también están presentes en las uvas. El calcio es un constituyente importante de la pared celular y de las pectinas, mientras que el magnesio forma parte de la clorofila y de las membranas celulares, además de participar como catalizador de enzimas implicados en la respiración, fotosíntesis y síntesis de ácidos nucleicos. De gran importancia también es el sodio por su contribución a la bomba de sodio-potasio, en las mitocondrias o en el movimiento de la savia por el floema. Los minerales se encuentran principalmente en las partes sólidas de las uvas, la piel (1 - 2%), las paredes celulares de la pulpa (0,8 – 2,8%), semillas (2 - 4%) y partes leñosas (2 - 3 %) [16,20].

A modo de resumen de la composición química de las uvas, se presenta la tabla 1.3 con los componentes principales en % peso fresco.

Tabla 1.3. Componentes químicos principales en uvas (% peso fresco) [21].

Compuestos	Rango
<i>Agua</i>	70–85
<i>Carbohidratos</i>	15–25
Glucosa	8–13
Fructosa	7–12
Pentosas	0,01–0,05
Pectinas	0,01–0,1

<i>Ácidos Orgánicos</i>	0,3–1,5
Tartárico	0,2–1,0
Málico	0,1–0,8
Cítrico	0,01–0,05
Acético	0,00–0,02
<i>Compuestos Fenólicos</i>	0,01–0,15
Antocianinas	0,00–0,05
Taninos	0,01–0,1
<i>Compuestos Nitrogenados</i>	0,03–0,17
Aminoácidos	0,01–0,08
Amoníaco	0,001–0,012
<i>Minerales</i>	0,3–0,5

1.1.4.2. Jugo de uva

1.1.4.2.1. Características generales

La historia de los jugos o zumos de frutas se inscribe en el marco más amplio de la evolución humana, de la historia de la cultura, de la alimentación y, especialmente, de la historia de la ciencia y de la tecnología. En la actualidad, los jugos de frutas han alcanzado un estatus importante en el menú diario de personas pertenecientes a diferentes edades, clases, grupos y áreas debido a sus excepcionales cualidades nutricionales, funcionales y terapéuticas. El jugo de uva es una bebida natural y refrescante, y su peculiar sabor y valor nutricional ha llevado a una mayor popularidad y demanda de esta bebida. Este jugo presenta una cantidad relativamente elevada de compuestos bioactivos, principalmente fenólicos, que muestran propiedades antioxidantes, cardioprotectoras y antiinflamatorias [22,23].

El Código Alimentario Argentino (CAA) en su capítulo XII, artículo 1062 define al jugo de uva como “el producto obtenido exclusivamente por expresión de las distintas variedades de uvas o sus mezclas, sanas, limpias y maduras; filtrado y estabilizado. Deberá estar libre de semillas, restos de piel, así como de cualquier cuerpo o sustancia extraños. El color, olor y sabor serán los característicos de la fruta madura, sin olores o sabores extraños, y podrá expendirse gasificado con gas carbónico, a una presión no menor de 1,5 atm a 21°C. Presentará una acidez de 0,45 a 0,85/100 ml en ácido tartárico; un pH variable entre 3,1 y 4,1;

un contenido en azúcar expresado en azúcar invertido entre 17 y 25 g/100 ml". A su vez define (capítulo X, artículo 775bis) como jugo concentrado de uva "al producto obtenido del mosto de uva sin fermentar por deshidratación parcial mediante procesos térmicos al vacío o a presión normal o cualquier otro proceso físico, sin haber sufrido una caramelización sensible" [24].

Por su parte la Ley N° 14.878 "General de Vinos", en su artículo 17, inciso f) productos analcohólicos de la uva, aporta también sus definiciones de jugo de uva y mostos [25]:

- jugo de uva: producto de la molienda o prensado de la uva fresca, filtrado y estabilizado con productos aprobados antes de iniciarse el proceso de fermentación alcohólica. Se tolerará alcohol proveniente de fermentaciones accidentales, con un límite máximo de 1% en volumen.
- mosto virgen de uva: el proveniente de la molienda o prensado de la uva fresca en tanto no haya empezado a fermentar.
- mosto de uva en fermentación: aquel en proceso de fermentación, cuya riqueza alcohólica no exceda de 5% en volumen.
- mosto sulfitado: el mosto estabilizado con el agregado de anhídrido sulfuroso en dosis que establezca la legislación.
- mosto concentrado: el obtenido del mosto de la uva en sus diversos grados de concentración mediante procesos térmicos al vacío o al aire libre, sin haber sufrido caramelización sensible.

En otras palabras el jugo de uva es una bebida derivada del género *Vitis sp*, principalmente de las especies *V. labrusca*, *V. vinifera* y *V. rotundifolia*, susceptible de fermentación, pero no fermentado, obtenido a partir de la parte comestible de frutas sanas y maduras, frescas o conservadas por el frío o congelación, de una o varias especies mezcladas, que posea el color, el aroma y el sabor característicos del jugo de la fruta de la que proviene [26].

Los términos de jugo de uva y mosto no están claramente diferenciados en la literatura y generalmente se usan como sinónimos, la mayoría coincide que en el mosto es el jugo de uva

que contiene diversos elementos de la uva como puede ser la piel, las semillas, etc, siendo el punto de partida para la elaboración del vino.

Con respecto al mercado mundial, es el mosto concentrado el que tiene un papel preponderante, ya que es un ingrediente versátil utilizado por diversas industrias para la elaboración de jugos mezclas, golosinas, dulces, mermeladas, helados, jaleas, galletitas, pan, como edulcorante de bebidas gaseosas y en la industria farmacéutica. Argentina está entre los primeros exportadores de mosto del mundo, exportando el 90% de su producción [27]. La venta de mosto argentino se concentra hoy en cuatro países y representa más del 80% del total exportado. EEUU es el destino más importante, seguido por Japón, Sudáfrica y Canadá.

En la industria nacional, el mosto argentino se concentra en unas 15 empresas, algunas con dedicación exclusiva y otras diversificadas con elaboración y fraccionamiento de vinos. El total comercializado se reparte entre Mendoza y San Juan prácticamente en partes similares.

Con respecto a la calidad y autenticidad de los jugos, el control de la falsificación de la composición del jugo no solo afecta la calidad de los jugos de frutas, sino que también engaña a los consumidores que pagan más por un producto con declaraciones falsas en su rótulo. Es difícil estimar la cantidad de jugos de frutas adulterados en el mercado mundial de alimentos, pero no hay duda de que con una mayor globalización, el impacto de eventos únicos de adulteraciones alimentarias afectará a una población más grande y más amplia que nunca. Las falsificaciones de jugos de frutas implican una adulteración simple al agregar agua, azúcar y ácidos orgánicos, o enfoques más complicados al usar productos de frutas, tales como lavado de pulpa, cáscara y otros componentes de la fruta [28].

1.1.4.2.2. Composición química

La química de las uvas y los jugos depende de muchos factores, como la calidad y composición del suelo, el grado de madurez de las bayas, el estrés hídrico, la exposición a la luz solar del racimo, la patogénesis, el tipo de sistema de cultivo, los procedimientos agronómicos, el procesamiento previo de las bayas, el método de prensado, entre otros. Con todos estos

factores, las diferencias en la composición química parecen ser obvias para las uvas, jugos y vinos. El jugo de uva es un medio acuoso rico en azúcar que contiene diferentes clases de compuestos con concentraciones que van desde mg/L (metionina, lisina y trans-resveratrol) hasta g/L (azúcares y ácidos orgánicos). En este sentido, los jugos puros presentan 70-80 g/100 ml de agua, 15-25 g/100 ml de carbohidratos (sólidos solubles totales entre 14 y 20º Brix), en los que generalmente están presentes glucosa y fructosa. Además de los azúcares, los ácidos orgánicos son los compuestos más abundantes presentes en el jugo de uva, de los cuales, los ácidos tartárico, málico, cítrico, láctico, acético y succínico son los principales, siendo el ácido málico y tartárico los mayoritarios con una concentración de hasta el 90% de los ácidos totales presentes. La composición de ácidos y azúcares de las uvas está muy influenciada por muchos factores, como la variedad de uva, la química del suelo, el índice de madurez, la región climática, las prácticas culturales y la exposición a la luz solar. Adicionalmente, los jugos de uva contienen casi 0,20 g/100 ml de fibras dietéticas y 0,20-0,60 g/100 ml de contenido de cenizas, mientras que los contenidos de lípidos y nitrógeno son insignificantes y apenas tienen interés nutricional en la dieta humana. El pH de los jugos de uva generalmente es inferior a 4,0.

Ciertos grupos de componentes minoritarios están presentes en los jugos de uva, ellos abarcan el grupo de los elementos inorgánicos de la cenizas, en los que encontramos minerales como Ca, Mg, K, Na, P, Fe, Zn, Se, Co y Cu [22,29]. El contenido de minerales varía considerablemente debido a la variedad de uva, el clima, el suelo, la fertilización y otras prácticas culturales. Las uvas contienen varios compuestos nitrogenados, que incluyen cationes de amonio, algunos aminoácidos, péptidos y proteínas. La concentración total de nitrógeno de la fruta aumenta durante el período de maduración y los principales compuestos de nitrógeno que se encuentran en los jugos de uva son la arginina y la prolina como los principales aminoácidos, seguidos de la alanina, ácido g-amino butírico, cistationina, ácido glutámico, triptófano, metionina, serina, leucina, histidina y ácidos aspárticos. La concentración total de aminoácidos en el jugo de uva está comprendido entre 700 y 1500

mg/L. En relación con la concentración de vitaminas, el jugo de uva contiene un nivel medio de tiamina, riboflavina, piridoxina, α -tocoferol, colina, ácido fólico, niacina y ácido ascórbico.

Cuando se trata de jugos de uva, los compuestos fenólicos son la clase principal y más importante de compuestos bioactivos que están asociados con los beneficios fisiológicos en los humanos. Después de los azúcares y ácidos, los compuestos fenólicos son los constituyentes más abundantes presentes en las uvas, y representan entre 700 y 3000 mg/L de la concentración fenólica total, dependiendo del índice de madurez, la variedad, el origen geográfico y el tipo de suelo, la exposición a la luz solar y también el método utilizado para la cuantificación. Los compuestos fenólicos son productos metabólicos secundarios de plantas derivadas de la fenilalanina. En general, los jugos de uva contienen flavonoides (antocianinas, flavonoles, flavanoles, proantocianidinas) y no flavonoides (ácidos cinámico y benzóico y estilbenos). Los compuestos fenólicos se concentran principalmente en las pieles y semillas. La cantidad de estos compuestos tanto en las uvas como en los jugos varía considerablemente de acuerdo con algunos factores intrínsecos y extrínsecos, como el clima, la cantidad de exposición a la luz solar y el estrés hídrico, el estado nutricional y la patogénesis de la vid, la naturaleza y la composición química del suelo, altitud, variedad de uva, grado de maduración de la uva y pH [22].

1.1.4.2.3. Proceso de elaboración

El proceso para la producción de jugo de uva comienza con la cosecha de las uvas en la etapa de su madurez técnica (14-25º Brix de azúcares solubles totales), debiéndose procesar dentro de las 48 h posteriores a la cosecha para evitar la contaminación con bacterias y hongos y la producción de alcohol etílico por proceso fermentativo. En la molienda se realiza la separación del escobajo del grano de uva y luego se produce el estrujado de estos últimos. En esta etapa es opcional el agregado de enzimas pectolíticas con las siguientes ventajas: mejorar el escurrido, acelerar los fenómenos de clarificación, y para facilitar la maceración (trituration) de las membranas en la piel de la uva (acelerando la liberación de pigmentos, como

antocianinas y otros compuestos químicos en el jugo) y también para reducir el contenido de coloides (pectina), facilitando así el proceso de filtración para obtener el jugo de uva. Existen varios métodos para el procesamiento comercial de los jugos de uva, como el prensado en frío y el prensado en caliente. El método de prensado continuo en caliente es la técnica más utilizada, ya que el método de prensado en frío produce un rendimiento de jugo muy bajo (aproximadamente de 18%). El escurrido y/o prensado tiene por fin separar las partes sólidas del mosto.



Figura 1.6. Elaboración de jugo de uva.

Con la utilización de enzimas y temperaturas superiores a 60°C, el jugo de uva obtenido por prensado en caliente contiene sólidos totales más altos y una mayor concentración de compuestos fenólicos (incluidas las antocianinas) en comparación con los otros métodos. Obviamente, la calidad del jugo se ve significativamente afectada por el tiempo y la temperatura de extracción, así como por la concentración y el tipo de enzima empleada en el proceso. Después de la trituración, se elimina aproximadamente el 35% del jugo de

funcionamiento libre utilizando una pantalla de malla 40 y la pulpa restante se vacía en una prensa de tornillo continua. El jugo libre y el jugo de prensado se combinan y filtran al vacío para eliminar los sólidos insolubles en suspensión, produciendo jugo y orujo que contiene pieles y semillas. Finalmente, el jugo se pasa a través de un intercambiador de calor y las botellas esterilizadas se llenan y luego se pasteurizan, se enfrían, se etiquetan y se almacenan hasta su distribución [22,29].

1.1.4.3. Vino

1.1.4.3.1. Características generales

El vino (del latín *vinum*) es la bebida resultante de la fermentación alcohólica total o parcial de las uvas frescas (fruto de la especie *Vitis vinifera*), ya sea trituradas o no, o de mosto de uva, que resulta en una concentración real de alcohol no menor del 8,5% vol [30].

Según la Ley Nacional de Vinos 14.878, se considera “Vinos genuinos a los obtenidos por la fermentación alcohólica de la uva fresca y madura o del mosto de la uva fresca, elaborados dentro de la misma zona de producción. A este efecto, la reglamentación fijará los grados baumé mínimos de las uvas, según las zonas y las condiciones climáticas” [25].

Esta bebida tiene una gran importancia social y económica, ya que la producción de vino es una actividad agrícola relevante en la economía mundial [31,32]. Aunque la producción de vino se ha mantenido estable en los últimos años con un aumento en el año 2018, la tendencia del consumo mundial de vino ha aumentado ligeramente [33].

Desde un punto de vista químico, el vino es una matriz realmente compleja en constante evolución, que resulta de un largo proceso de transformaciones físicas, químicas y biológicas y compuesta de moléculas de diferente naturaleza y estructura (por ejemplo, compuestos volátiles como el etanol, proteínas, aminoácidos, carbohidratos, compuestos fenólicos e inorgánicos) [34]. Esta compleja bebida está influenciada por varios factores como el clima, suelo donde maduraron las uvas, la selección de variedades de uva, el momento y método de la cosecha, el proceso de fermentación, las condiciones de almacenamiento y

envejecimiento, así como todas las prácticas vitivinícolas en general [35]. Estas características hacen del vino una bebida única y valiosa con un valor agregado distintivo para el consumidor y el productor, basado en su origen geográfico y, por lo tanto, en su "terroir" asociado [34].

En el mundo de la vitivinicultura, el término francés "terroir" (de la palabra *terre*, que significa suelo) engloba una serie de factores que interaccionan en la elaboración del vino, este concepto será explicado más adelante en el punto 1.2.3. En este contexto, surgen los sistemas de indicación geográfica (IG) como indicadores de autenticidad y calidad del vino que se establecen en la mayoría de los países productores de vino [36]. Existe entonces un estricto marco legal para la trazabilidad del vino y el control de calidad con el fin de garantizar la seguridad y la excelencia de esta bebida. Desde el punto de vista industrial, la trazabilidad también es un parámetro útil para controlar y monitorear los procesos de vinificación [34,37].

La verificación y evaluación actual de la autenticidad, calidad, y las adulteraciones fraudulentas del vino requieren evaluar y verificar tres factores clave como la declaración de origen, la composición real de acuerdo con la etiqueta (variedad de uva y producción enológica) y la cosecha [38]. Además, como en todas las áreas de producción de alimentos y bebidas, se requieren mediciones analíticas precisas en todas las etapas del proceso de elaboración del vino, comenzando en el viñedo y terminando en los pasos finales de embotellado y con la certificación para garantizar el control de calidad del producto.

El fraude al vino es un problema que está presente en varias formas; en todos los casos, está relacionado con la calidad y los aspectos comerciales del vino. El tipo de fraude más extendido es la adulteración de esta bebida con la adición de sustancias que no están relacionadas con las uvas, principalmente con productos de bajo costo y, a veces, con productos químicos peligrosos, por ejemplo, la dilución con agua o adición de alcohol [34,39]. La falsificación y la información errónea sobre el vino, como el etiquetado incorrecto, es otro tipo común de fraude al vino [39,40].

La adulteración del vino, principalmente en términos de variedades y regiones de origen, ha sido motivo de gran preocupación y aún hoy en día lo sigue siendo [37,41]. La legislación vitivinícola relacionada a la indicación de origen es completa y difiere según el país e incluso la región, que está estrictamente regulada por organizaciones internacionales o agencias gubernamentales. En materia de etiquetado, diferentes países poseen diferentes nombres, como "Denominazione di Origine Controllata" de Italia, "Institut National des Appellations d'Origine" de Francia, "Denominación de Origen" de España, "Wine of Origin" de Sudáfrica, o las "American Viticultural Areas" de Estados Unidos, para la autenticidad del vino [32,42]. Para evaluar la calidad del vino estrictamente vinculada a un área geográfica particular o métodos específicos de producción y para proteger a los consumidores, la Comisión Europea ha definido las designaciones: Denominación de Origen Protegida (PDO, sigla en inglés) e Indicación Geográfica Protegida (PGI, sigla en inglés) y Garantía de Especialidad Tradicional (TSG, sigla en inglés) detalladas en el punto 1.2.2 de la presente tesis doctoral.

En este sentido, la existencia de diversas técnicas analíticas ha sido útil para clasificar las muestras de vino de acuerdo con su terruño, origen geográfico, variedad y añejado. Asimismo, los métodos quimiométricos tienen como objetivo identificar marcadores únicos o grupos de marcadores que caracterizan la autenticidad.

1.1.4.3.2. Composición química

Como se dijo anteriormente, el vino es una matriz compleja, en la cual coexisten ciertas sustancias que provienen de la uva y en el mosto y otras que se forman durante la fermentación alcohólica y maloláctica, además de sustancias originadas en las múltiples reacciones químicas y biológicas que tienen lugar durante la conservación y crianza.

Para comprender lo que es el vino desde el punto de vista de sus componentes, hay que comprender los cambios de composición que se producen al evolucionar desde la uva al mosto y al vino. El mosto antes de la fermentación se compone principalmente de agua y azúcares, así como ácidos (tartárico y málico) y componentes minoritarios. En el vino, además,

se incorporan nuevos componentes químicos en menor o mayor cantidad, responsables de la composición final.

El agua es el compuesto químico más abundante del vino [43]. Este componente es esencial para el desarrollo de muchas reacciones químicas que ocurren durante la fermentación y el envejecimiento del vino. El etanol es el segundo componente principal del vino, seguido de glicerol, ácidos orgánicos, compuestos fenólicos, vitaminas y minerales.

Alcoholes

El etanol es el alcohol más abundante presente en el vino y se produce durante la fermentación alcohólica realizada por las levaduras. La concentración de etanol es un parámetro importante a considerar en el procesamiento del vino, ya que tiene un efecto directo sobre las propiedades organolépticas y evita el crecimiento de microorganismos patógenos en esta bebida [18]. Aunque existen límites legales para la concentración de alcohol en vinos, fijados por diferentes organismos de control nacionales e internacionales, es muy preocupante que en los últimos años la graduación alcohólica de los vinos está creciendo debido al aumento de la temperatura climática causada por calentamiento global [44]. En climas cálidos, las uvas contienen una mayor concentración de azúcares y, por lo tanto, producen vinos con mayores concentraciones de etanol, incluso por encima del 15% (v/v).

Aunque el etanol es el compuesto representativo dentro del grupo de alcoholes, también se puede producir una cantidad mínima de otros alcoholes que incluyen propanol, butanol y glicerol durante la fermentación alcohólica. Este último alcohol confiere viscosidad y dulzura a los vinos; sin embargo, una alta concentración de glicerol puede ser indicativa de podredumbre noble causada por cepas de *Botrytis* [45,46].

Ácidos orgánicos

Los ácidos orgánicos están presentes en el vino y contribuyen con sus propiedades organolépticas, por ejemplo, el color del vino tinto. Estos compuestos también colaboran con la conservación del vino, mejorando la estabilidad fisicoquímica y microbiológica de la bebida

[47]. Algunos de los ácidos orgánicos que se encuentran en el vino provienen directamente de las uvas, como el tartárico, L (-) -málico, cítrico y ácido ascórbico. Durante la fermentación, se generan otros ácidos orgánicos, incluidos pirúvico, L (+) - láctico, D(-) - láctico, succínico, acético, citramalico, oxaloacético y fumárico [47]. Los ácidos orgánicos a menudo están involucrados en reacciones químicas para formar ésteres, que son parcialmente responsables de la fragancia a fruta fresca de los vinos [18].

Carbohidratos

Con respecto a los azúcares, glucosa y fructosa se consideran los precursores directos del etanol porque son abundantes en las uvas y son azúcares fermentables. Por lo tanto, después de la fermentación alcohólica, la concentración de glucosa y fructosa disminuye considerablemente. Dado que la mayoría de las cepas de levadura prefieren la glucosa, se consume en un porcentaje mayor que la fructosa [48]. También se han identificado varios disacáridos en los vinos, incluyendo sacarosa, maltosa, lactosa, melibiosa, rafinosa y trehalosa [49,50]. Las pectinas son carbohidratos más complejos que también pueden estar presentes en los vinos, aunque a bajas concentraciones.

Compuestos fenólicos

Los vinos contienen una amplia variedad de compuestos fenólicos, incluidos flavonoides, flavonoles, antocianinas, ácidos fenólicos, estilbenos entre otros. Los polifenoles tienen diferentes propiedades, pero las más interesantes desde la perspectiva vitivinícola son:

- Actividad bactericida y antifúngica.
- Propiedades sensoriales: participa en el color del vino, sabor, sensación astringente y aroma de los fenoles volátiles.

Los compuestos fenólicos están presentes en el mosto, y sus niveles aumentan considerablemente durante la vinificación. Estos compuestos juegan un papel importante en los vinos tintos ya que estos se fermentan en las pieles de uva, que contienen las

concentraciones más altas de polifenoles. Las pieles de uva son particularmente ricas en antocianinas que confieren color [51].

Compuestos nitrogenados

Los aminoácidos, las proteínas y los péptidos representan los compuestos de nitrógeno más abundantes de mostos y vinos. Los aminoácidos son útiles en la etapa de fermentación alcohólica, ya que pueden servir como sustratos nutritivos para las levaduras [52]. Además, contribuyen a las características aromáticas de los vinos. La concentración de aminoácidos se modifica durante la producción de vino. Por ejemplo, se ha informado que después de la fermentación maloláctica, la concentración de 15 aminoácidos (serina, glicina, treonina, alanina, prolina, isoleucina, triptófano, leucina, lisina, fenilalanina, tirosina, valina, cisteína, ácido aspártico, gamma- ácido minobutírico) aumentó entre 4,6% y 35,2%, mientras que la concentración de glutamina, arginina, histidina y metionina disminuyó entre 4,36% y 29,1% [53].

La concentración de proteínas y péptidos puede aumentar en los pasos iniciales de la fermentación alcohólica, pero luego disminuye como resultado de la presencia de bentonitas y con el envejecimiento del vino. Estos compuestos tienen un impacto en las propiedades organolépticas de los vinos, ya que colaboran con las características aromáticas y gustativas de la bebida. La presencia de altas concentraciones de estos grupos de compuestos en los vinos generalmente se asocia con alergia al vino [54].

Vitaminas

Las vitaminas representan una fracción mínima de la composición del vino. La concentración de vitaminas disminuye notablemente durante la elaboración del vino, especialmente durante las etapas de fermentación y envejecimiento [18]. Las vitaminas juegan un papel esencial en la fermentación, ya que son factores de crecimiento para la levadura y las bacterias responsables de la fermentación alcohólica y la fermentación maloláctica, respectivamente. Aunque la tiamina desaparece casi por completo y los niveles de ácido

ascórbico y piridoxina disminuyen, el vino sigue siendo una valiosa fuente potencial de vitaminas [51].

Minerales

Finalmente con respecto a los minerales, la fermentación alcohólica exitosa depende, entre otras cosas, de la presencia de K, Mg, Cu, Fe, Ca, Co y Zn. Los minerales (aniones o cationes) están presentes en el mosto en forma de sales minerales (halogenuros, sulfatos, fosfatos, etc.) y orgánicas (tartrato, malato, lactato, etc.) y durante la vinificación sus concentraciones se modifican mediante la adición de bentonita y compuestos similares utilizados para clarificar el vino, mientras que algunos subsisten en su mayor parte en el vino, pero en proporciones más bajas en algunos casos, dadas las insolubilidades que se producen durante las fermentaciones y durante el envejecimiento del vino.

Los niveles de potasio y calcio generalmente disminuyen a medida que aumenta el contenido de etanol del medio. Los niveles de otros cationes metálicos como el hierro y el cobre pueden aumentar debido al contacto entre el mosto o el vino y el hierro, el acero inoxidable o el bronce [51].

Las concentraciones de los minerales están influenciadas por varios factores, como el suelo, el clima, la variedad de uva, la contaminación, las prácticas de viticultura (por ejemplo, la aplicación de fungicidas), el uso de bentonitas y todo el proceso utilizado para la producción de vino. La composición mineral en los vinos es importante ya que puede afectar la estabilidad de los vinos y las propiedades organolépticas. Además, debe estar de acuerdo con los límites legales, ya que algunos elementos son altamente tóxicos, como As, Hg, Pb y Cd, entre otros. Finalmente, el perfil mineral puede ser útil como herramienta para diferenciar los vinos por origen geográfico ya que la composición mineral está condicionada por el suelo en el que se cultivan las vides.

1.1.4.3.3. Proceso de elaboración

La elaboración del vino comienza con la cosecha de las uvas maduras, es decir de aquellas uvas que hayan llegado al máximo contenido de azúcar. La cosecha puede ser manual o mecánica. En la bodega la uva es depositada en el lagar o dispuesta en la cinta de selección, en la que se separan racimos o granos defectuosos, hojas o cualquier elemento extraño que pueda venir con la uva. Se procede al despallado o descobajado donde los granos de uva se separan del escobajo (parte verde del racimo). Posteriormente se procede a la molienda o estrujado donde la uva pasa por rodillos acanalados que los rompen suavemente para permitir un rápido escurrimiento del jugo durante el proceso de prensado. En el prensado la uva es sometida a una presión para extraer la mayor cantidad de jugo o mosto. La presión no debe ser excesiva para no extraer sustancias amargas y astringentes de las semillas y/o del hollejo. Se utilizan prensas neumáticas que trabajan con presión reducida, obteniéndose mostos de gran calidad. Una vez que el mosto y los orujos son colocados en la vasija, se agrega una dosis de anhídrido sulfuroso que cumple con dos funciones principales: antiséptico y antioxidante. Para elaborar el vino tinto el mosto se encuba junto con la pulpa, el hollejo y las semillas o pepitas. Se puede colocar en distintos tipos de vasija: acero inoxidable o piletas de hormigón armado. En el caso de los vinos blancos los mostos son llevados a una vasija donde permanecen de 8 a 24 horas a baja temperatura. Dado que los mostos están inmovilizados se produce una decantación de restos de pulpas, semillas, tierra y partículas en suspensión que sedimentan en el fondo de la vasija y se separan luego mediante un trasiego.

Los vinos tintos siguen con la fermentación alcohólica. Se realiza a temperaturas que van desde los 24°C a los 30°C dependiendo del tipo de vino a elaborar. Pasa lo mismo con los tiempos de encubado que varían entre 7 a 20 días o más. En esta etapa, además de comenzar la fermentación, se produce otro proceso que es el de maceración, por el cual el jugo de la uva al estar en contacto con la piel y otros elementos sólidos irá ganando color, taninos, aromas y sabores. En los tintos de guarda, la fermentación es a temperaturas más altas y las

maceraciones son más prolongadas ya que le otorgan una mayor estructura al vino, con más cantidad de polifenoles y mayor suavidad en los taninos. La fermentación se completa cuando la cantidad de azúcares reductores es inferior a los dos gramos por litro. En ese momento se procede al descube o la separación del vino flor de los orujos (semillas, hollejos y pulpa). Se prensan los orujos y se extrae el llamado vino prensa. Posteriormente el ácido málico es de suma importancia biológica para el vino. Durante la fermentación alcohólica el ácido málico comienza a transformarse por las levaduras y ciertas bacterias lácticas, en ácido láctico. Pero terminada la fermentación alcohólica estas bacterias que suceden a las levaduras alcohólicas efectúan lo que se conoce como fermentación maloláctica. En ella el ácido málico termina de transformarse en ácido láctico, anhídrido carbónico y una serie de compuestos, que aportan a los vinos aromas mantecosos. En esta etapa el vino debe mantenerse a temperaturas cercanas a 20°C. Sigue el trasiego esta operación consiste en sacar los vinos que se encuentran sobre borras y pasarlos a otra vasija limpia, evitando arrastrar los sedimentos. La clarificación consiste en agregar al vino una sustancia de naturaleza coloidal (mineral, vegetal o animal) que es capaz de arrastran hacia el fondo de la vasija aquellos elementos en suspensión no deseados en el vino. Los vinos tintos pueden ser jóvenes, frescos, frutados. Otros vinos pueden atravesar un proceso llamado crianza, que consiste en estacionar los vinos generalmente en barricas de madera (generalmente roble francés o americano), lo cual permitirá, después de la fermentación maloláctica, que el vino se enriquezca en aromas, sabores y estructura. La crianza en barricas, además favorece la estabilización del color. Cuando se completa la crianza se procede a fraccionar el vino y embotellarlo cuidadosamente. En la siguiente Figura 1.7 se presentan los principales pasos de la elaboración de vino.

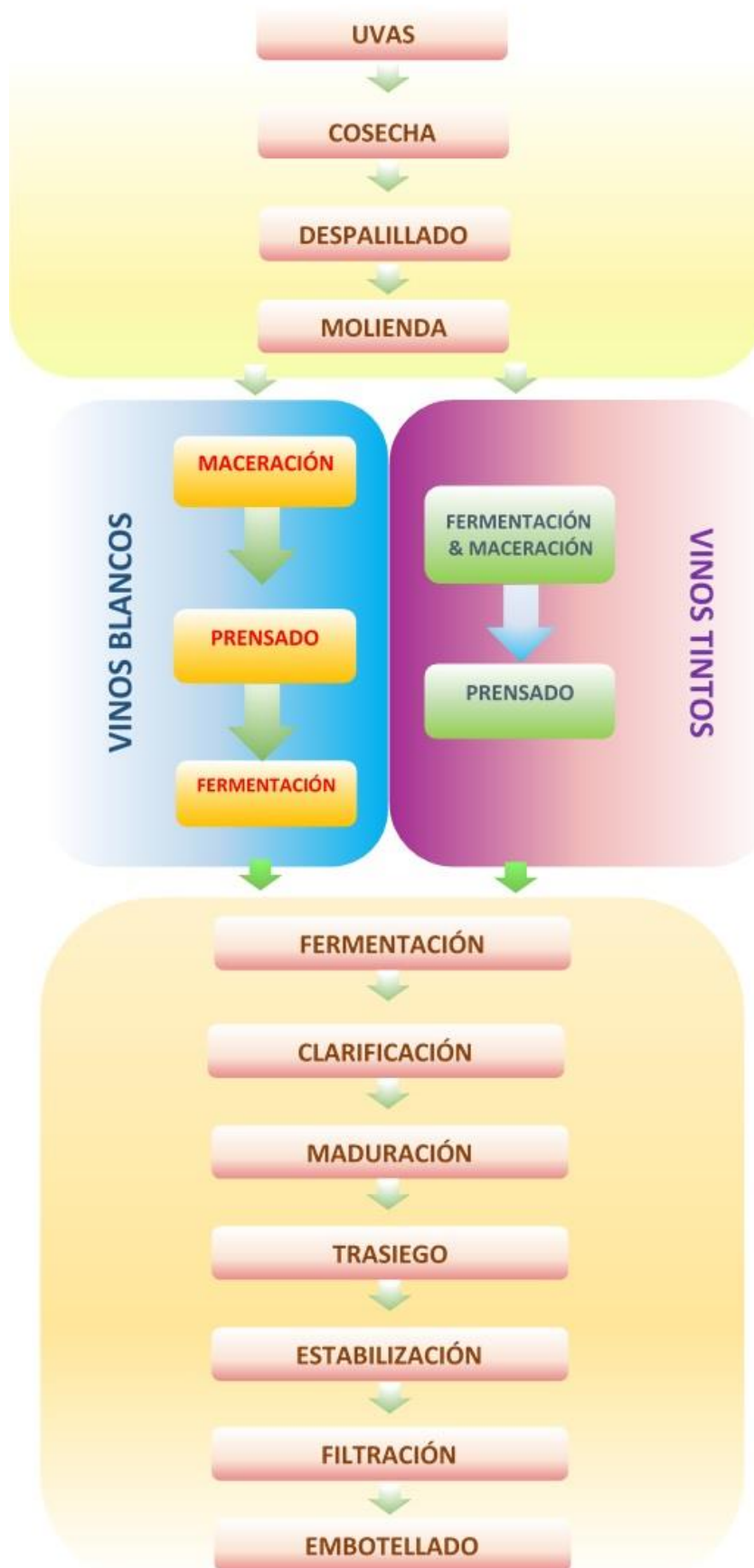


Figura 1.7. Principales pasos de la elaboración de vino tinto y blanco.

1.1.5. Referencias

- [1] R.S. Jackson, Wine Science. Principles and Applications, Third, Elsevier Inc., 2008.
<https://www.agronomy.org/publications/jeq/abstracts/25/3/JEQ0250030629a>.
- [2] P.E. McGovern, Neolithic Wine!, in: Anc. Wine, Princeton University Press, 2019: pp. 64–84. doi:10.2307/j.ctvfjd0bk.9.
- [3] J. Fehér, G. Lengyel, A. Lugasi, The cultural history of wine - Theoretical background to wine therapy, Cent. Eur. J. Med. 2 (2007) 379–391. doi:10.2478/s11536-007-0048-9.
- [4] J.L. Lanzarini, J. Mangione, La cultura de la vid y el vino: La vitivinicultura hace escuela, 1ª, Fondo Vitivinícola Mendoza, Mendoza, 2009.
- [5] A. Yann Guisard, John Blackman, Andrew Clark, Bruno Holzapfel, J.S. and C.S. Rawson, Suzy Rogiers, Leigh Schmidtke, Horticulture: Plants for People and Places, Volume 1, Springer Netherlands, Dordrecht, 2014. doi:10.1007/978-94-017-8578-5.
- [6] H. Li, H. Wang, H. Li, S. Goodman, P. van der Lee, Z. Xu, A. Fortunato, P. Yang, The worlds of wine: Old, new and ancient, Wine Econ. Policy. 7 (2018) 178–182. doi:10.1016/j.wep.2018.10.002.
- [7] Dirección General De Escuelas, La Vitivinicultura en Mendoza, (2016).
<http://www.mendoza.edu.ar/la-vitivinicultura-en-mendoza/> (accessed August 14, 2019).
- [8] Wines of Argentina, Mendoza, capital mundial del vino, (2013).
<http://www.winesofargentina.org/> (accessed August 11, 2019).
- [9] Instituto Nacional de Vitivinicultura (INV), Informe Anual De Superficie 2018, (2019).
www.inv.gob.ar (accessed October 10, 2019).
- [10] C.D. Catania, S. Avagnina de del Monte, R.F. del Monte, J. Tonietto, M.C. Zanus, Caracterização climática de regiões vitivinícolas Ibero-Americanas, 1º, Embrapa Uva e Vinho, Bento Gonçalves, Brasil, 2007.
- [11] Subgerencia de Estadísticas y Asuntos Técnicos Internacionales, Instituto Nacional de

- Vitivinicultura., Regiones Vitivinícolas Argentinas. Provincia de Mendoza. Área Sur, 2019. www.inv.gob.ar.
- [12] International Organization of Vine and Wine (OIV), State of the sector in 2018, 2019. <http://www.oiv.int/public/medias/6679/en-oiv-state-of-the-vitiviniculture-world-market-2019.pdf>.
- [13] Instituto Nacional de Vitivinicultura (INV), Informe anual de cosecha y elaboración 2018, (2018). www.inv.gob.ar (accessed October 10, 2019).
- [14] N. Fontes, H. Gerós, S. Delrot, Grape Berry Vacuole: A Complex and Heterogeneous Membrane System Specialized in the Accumulation of Solutes, *Am. J. Enol. Vitic.* 62 (2011) 270–278. doi:10.5344/ajev.2011.10125.
- [15] F. Cosme, B. Gonçalves, A. Ines, A.M. Jordão, A. Vilela, Grape and Wine Metabolites: Biotechnological Approaches to Improve Wine Quality, in: *Grape Wine Biotechnol., InTech*, 2016: p. 13. doi:10.5772/64822.
- [16] C. Conde, P. Silva, N. Fontes, A.C.P. Dias, R.M. Tavares, M.J. Sousa, A. Agasse, S. Delrot, H. Gerós, Biochemical changes throughout grape berry development and fruit and wine quality, *Food. 1* (2007) 1–22. <http://hdl.handle.net/1822/6820>.
- [17] P. Diakou-Verdin, J.-P. Carde, J.-P. Gaudillère, F. Barrieu, N. Ollat, A. Moing, Grape berry development : A review, *OENO One.* 36 (2002) 109. doi:10.20870/oeno-one.2002.36.3.970.
- [18] R.S. Jackson, 6 - Chemical Constituents of Grapes and Wine BT - Wine Science (Fourth Edition), in: *Food Sci. Technol., Academic Press, San Diego*, 2014: pp. 347–426. doi:<http://dx.doi.org/10.1016/B978-0-12-381468-5.00006-3>.
- [19] M.P. Fabani, M.E. Toro, F. Vázquez, M.P. Díaz, D.A. Wunderlin, Differential absorption of metals from soil to diverse vine varieties from the valley of tulum (Argentina): Consequences to evaluate wine provenance, *J. Agric. Food Chem.* 57 (2009) 7409–7416. doi:10.1021/jf901572k.

- [20] M.P. Fabani, M.E. Toro, F. Vázquez, M.P. Díaz, D.A. Wunderlin, Differential Absorption of Metals from Soil to Diverse Vine Varieties from the Valley of Tulum (Argentina): Consequences To Evaluate Wine Provenance, *J. Agric. Food Chem.* 57 (2009) 7409–7416. doi:10.1021/jf901572k.
- [21] M.A. Amerine, H.W. Berg, R.E. Kunkee, C.S. Ough, V.L. Singleton, A.D. Webb, *The Technology of Wine Making*, 4th ed., AVI Publishing Company, Inc., Westport, Connecticut, 1980.
- [22] D. Granato, M. de Magalhães Carrapeiro, V. Fogliano, S.M. van Ruth, Effects of geographical origin, varietal and farming system on the chemical composition and functional properties of purple grape juices: A review, *Trends Food Sci. Technol.* 52 (2016) 31–48. doi:10.1016/j.tifs.2016.03.013.
- [23] R.L. Bhardwaj, U. Nandal, A. Pal, S. Jain, Bioactive compounds and medicinal properties of fruit juices, *Fruits.* 69 (2014) 391–412. doi:10.1051/fruits/2014027.
- [24] Secretaría de Estado de Salud Pública, Código Alimentario Argentino. Ley 18284., Artículos 155 Al 183 - Los Prod. Aliment. (2017).
<https://www.argentina.gob.ar/anmat/codigoalimentario> (accessed March 4, 2019).
- [25] Instituto Nacional de Vitivinicultura, Ley N° 14.878. Ley General de Vinos, Argentina, 1959.
- [26] K. Mihalev, R. Dinkova, V. Shikov, P. Mollov, Classification of Fruit Juices, in: *Fruit Juices*, Elsevier, 2018: pp. 33–44. doi:10.1016/B978-0-12-802230-6.00003-5.
- [27] Cámara Argentina de Fabricantes y Exportadores de Mosto, El 90% del jugo de uva argentino está destinado a la exportación, (2019).
<http://www.prensa.mendoza.gov.ar/el-90-del-jugo-de-uva-argentino-esta-destinado-a-la-exportacion/> (accessed July 15, 2019).
- [28] P.R. Ashurst, Fruit Juices, in: *Encycl. Food Heal.*, Elsevier, 2016: pp. 130–137. doi:10.1016/B978-0-12-384947-2.00334-2.

- [29] I.M. Toaldo, F.A. Cruz, T.D.L. Alves, J.S. de Gois, D.L.G. Borges, H.P. Cunha, E.L. da Silva, M.T. Bordignon-Luiz, Bioactive potential of *Vitis labrusca* L. grape juices from the Southern Region of Brazil: Phenolic and elemental composition and effect on lipid peroxidation in healthy subjects, *Food Chem.* 173 (2015) 527–535. doi:10.1016/j.foodchem.2014.09.171.
- [30] International Organization of Vine and Wine (OIV), International Code of Oenological Practices, 33 (2016) 1–29. <http://www.oiv.int/en/technical-standards-and-documents/oenological-practices/international-code-of-oenological-practices> (accessed September 29, 2017).
- [31] C. Villano, M.T. Lisanti, A. Gambuti, R. Vecchio, L. Moio, L. Frusciante, R. Aversano, D. Carputo, Wine varietal authentication based on phenolics, volatiles and DNA markers: State of the art, perspectives and drawbacks, *Food Control.* 80 (2017) 1–10. doi:10.1016/j.foodcont.2017.04.020.
- [32] G. Grindlay, J. Mora, L. Gras, M.T.C. de Loos-Vollebregt, Atomic spectrometry methods for wine analysis: A critical evaluation and discussion of recent applications, *Anal. Chim. Acta.* 691 (2011) 18–32. doi:10.1016/j.aca.2011.02.050.
- [33] (OIV) International Organization of Vine and Wine, 2019 Statistical Report on World Vitiviniculture, (2019) 1–23. <http://www.oiv.int/public/medias/6759/oiv-2019-statistical-report-on-world-vitiviniculture.pdf> (accessed July 16, 2019).
- [34] M.E. Alañón, M.S. Pérez-Coello, M.L. Marina, Wine science in the metabolomics era, *TrAC Trends Anal. Chem.* 74 (2015) 1–20. doi:10.1016/j.trac.2015.05.006.
- [35] C.A. Teixeira dos Santos, R.N.M.J. Páscoa, M.C. Sarraguça, P.A.L.S. Porto, A.L. Cerdeira, J.M. González-Sáiz, C. Pizarro, J.A. Lopes, Merging vibrational spectroscopic data for wine classification according to the geographic origin, *Food Res. Int.* (2017). doi:10.1016/j.foodres.2017.09.018.
- [36] C.A.T. dos Santos, R.N.M.J. Páscoa, J.A. Lopes, A review on the application of vibrational

- spectroscopy in the wine industry: From soil to bottle, *TrAC - Trends Anal. Chem.* 88 (2017) 100–118. doi:10.1016/j.trac.2016.12.012.
- [37] A. Versari, V.F. Laurie, A. Ricci, L. Laghi, G.P. Parpinello, Progress in authentication, typification and traceability of grapes and wines by chemometric approaches, *Food Res. Int.* 60 (2014) 2–18. doi:10.1016/j.foodres.2014.02.007.
- [38] J. Saurina, Characterization of wines using compositional profiles and chemometrics, *TrAC Trends Anal. Chem.* 29 (2010) 234–245. doi:10.1016/j.trac.2009.11.008.
- [39] L. Holmberg, Wine fraud, *Int. J. Wine Res.* 2 (2010) 105–113. doi:10.2147/IJWR.S14102.
- [40] N. Dordevic, Chemometrics and stable isotope ratios of wine, 2015. <http://repository.ubn.ru.nl/handle/2066/147420>.
- [41] I.S. Arvanitoyannis, Wine authenticity, traceability and safety monitoring, Woodhead Publishing Limited, 2010. doi:<http://dx.doi.org/10.1533/9781845699284.2.218>.
- [42] G.P. Danezis, A.S. Tsagkaris, F. Camin, V. Brusic, C.A. Georgiou, Food authentication: Techniques, trends & emerging approaches, *TrAC - Trends Anal. Chem.* 85 (2016) 123–132. doi:10.1016/j.trac.2016.02.026.
- [43] R.W. Hutkins, Wine Fermentation, in: *Microbiol. Technol. Fermented Foods*, First edit, Blackwell Publishing, Ames, Iowa, USA, 2006: pp. 349–395. doi:10.1002/9780470277515.ch10.
- [44] B. Ozturk, E. Anli, Different techniques for reducing alcohol levels in wine: A review , *BIO Web Conf.* 3 (2014) 1–8.
- [45] R. Gawel, E.J. Waters, The effect of glycerol on the perceived viscosity of dry white table wine, *J. Wine Res.* 19 (2008) 109–114. doi:10.1080/09571260802622191.
- [46] B. Zoecklein, K.C. Fugelsang, B.H. Gump, F.S. Nury, *Wine Analysis and Production*, Springer, New York, 1999.
- [47] J. Ribéreau-Gayon, Y. Glories, A. Maujean, D. Dubourdieu, *The Chemistry of Wine: Stabilization and Treatments*, 2nd ed., John Wiley & Sons, Ltd, France, 2006.

- [48] J. Tronchoni, A. Gamero, F.N. Arroyo-Lopez, E. Barrio, A. Querol, Differences in the glucose and fructose consumption profiles in diverse *Saccharomyces* wine species and their hybrids during grape juice fermentation, *Int J Food Microbiol.* 134 (2009) 237–243. doi:10.1016/j.ijfoodmicro.2009.07.004.
- [49] A. Bertrand, M. O Dubernet, P. Ribéreau-Gayon, Trehalose, the principle disaccharide of wine, *Comptes Rendus Hebdomadaires des Seances de L'Academie des Sciences*, France, 1975.
- [50] A.I. Ruiz-Matute, M.L. Sanz, M. V Moreno-Arribas, I. Martinez-Castro, Identification of free disaccharides and other glycosides in wine, *J Chromatogr A.* 1216 (2009) 7296–7300. doi:10.1016/j.chroma.2009.08.086.
- [51] J. Moreno, R. Peinado, Composition of Wine, in: J. Moreno, R. Peinado (Eds.), *Enol. Chem.*, Elsevier, 2012: pp. 41–52. doi:10.1016/B978-0-12-388438-1.00004-2.
- [52] M. V Moreno-Arribas, C. Polo, *Wine Chemistry and Biochemistry*, Springer New York, New York, NY, 2009. doi:10.1007/978-0-387-74118-5.
- [53] P.M. Izquierdo Cañas, E. García Romero, S. Gómez Alonso, M. Fernández González, M.L.L. Palop Herreros, Amino acids and biogenic amines during spontaneous malolactic fermentation in Tempranillo red wines, *J. Food Compos. Anal.* 21 (2008) 731–735. doi:https://doi.org/10.1016/j.jfca.2007.11.002.
- [54] N. Jaeckels, I. Bellinghausen, P. Fronk, B. Heydenreich, J. Saloga, H. Decker, Assessment of sensitization to grape and wine allergens as possible causes of adverse reactions to wine: a pilot study, *Clin. Transl. Allergy.* 5 (2015) 21. doi:10.1186/s13601-015-0065-8.

1.2. DENOMINACIÓN DE ORIGEN Y TRAZABILIDAD

1.2.1. La importancia del origen de los alimentos

En las últimas décadas, la sociedad se ha ido volviendo cada vez más consciente de la importancia de la calidad de los alimentos. Los consumidores de todo el mundo exigen cada vez más productos seguros y de alta calidad, entendiendo la calidad como la suma de características y propiedades de un producto que influyen en su capacidad para satisfacer necesidades declaradas o implícitas. De acuerdo con la Organización Internacional de Normalización o Estandarización (ISO 9000:2005), la calidad se define como el grado en que un conjunto de características inherentes cumplen los requisitos [1]. Éste es un término flexible en el que los atributos intrínsecos como el sabor, la textura y la vida útil y los extrínsecos como el uso de pesticidas, el tipo de material de empaque, la tecnología de procesamiento, entre otros, son características importantes [2]. La autenticidad es una garantía adicional de la calidad y seguridad de un alimento, ya que verifica diferentes aspectos tales como la caracterización completa del producto, identificación de adulteraciones, control del cumplimiento de la etiqueta y evaluación del origen geográfico [3]. La verificación de la procedencia es un tema importante para la seguridad alimentaria, la calidad de los alimentos y la protección del consumidor, así como para el cumplimiento de la legislación nacional y las normas internacionales. Debido a la globalización de los mercados de alimentos y al aumento resultante en la variabilidad y disponibilidad de productos alimenticios de otros países, los consumidores están cada vez más interesados en conocer el origen geográfico junto con la calidad asumida de los productos que comen y beben. El aseguramiento de la calidad y los métodos utilizados para autenticar los alimentos son de gran interés tanto desde el punto de vista comercial como legal [4].

A su vez, el tema de la calidad vinculada al origen de los alimentos tiene particular relevancia para el desarrollo rural. Algunos productos agrícolas se diferencian por ciertas características de calidad o reputación debidas principalmente a su origen. Esta diferenciación

se puede atribuir a la tipicidad, la historia del producto, y a su carácter distintivo asociado a factores naturales o humanos, como el suelo, el clima, los conocimientos locales o las tradiciones. Durante miles de años, los consumidores han tenido un conocimiento íntimo del origen de sus alimentos, ya sea de cosecha propia o adquiridos en su productor/mercado local. Más recientemente, la preferencia del consumidor por el acceso continuo a una amplia variedad de alimentos frescos ha transformado las cadenas de suministro de locales a globales, disminuyendo así el conocimiento del consumidor sobre quién produjo sus alimentos y cómo. Como resultado de ello, los organismos de control y la industria alimentaria han introducido la regulación, los sistemas de trazabilidad y las estrategias de comercialización destinadas a tranquilizar al consumidor sobre la calidad, seguridad y procedencia, es decir, sobre la integridad de sus alimentos.

La trazabilidad implica la capacidad de identificar en cualquier etapa específica de la cadena alimentaria, el origen, los procesos de producción hasta la distribución de un producto, es decir es la capacidad de recrear la historia de un producto [5]. La trazabilidad permite la alerta temprana de problemas y peligros de calidad originando el retiro eficiente del producto cuando fuese necesario. Cada unidad comercial de cada material alimenticio recibe un identificador único que lo acompaña y se registra en todas las etapas de su progreso a través de su cadena alimentaria, permitiendo el monitoreo detallado de las propiedades de los productos alimenticios a medida que se elaboran desde la cosecha hasta el transporte, el almacenamiento, el procesamiento, la distribución y las ventas.

Existen varias motivaciones para desarrollar métodos que permitan verificar el origen de los alimentos. Una es garantizar la integridad de la cadena alimentaria y ayudar en la trazabilidad de los alimentos. Otra es el monitoreo y cumplimiento de los requisitos obligatorios de etiquetado de origen o verificación de reclamos, como parte de los programas de cumplimiento de normas alimentarias y la prevención del fraude alimentario [6]. El suministro de información sobre el origen de un alimento es relativamente reciente y, en

general, solo se aplica a ciertos alimentos específicos. A su vez, las razones para este interés en la procedencia van desde (a) patriotismo; (b) cualidades culinarias u organolépticas específicas asociadas con productos regionales; (c) disminución de la confianza en la calidad y seguridad de los productos producidos fuera de su región local, país, o de procedencia desconocida; o (d) preocupación por el bienestar animal y los métodos de producción percibidos como "ecológicos" y más a menudo adoptado por pequeños productores regionales. Desde el enfoque del agricultor, el resultado es que los productos con origen geográfico conocido se pueden vender a precios marcadamente más altos que los productos similares de otra procedencia o procedencia desconocida y, por lo tanto, mejora la competitividad y rentabilidad de los productores. Para ciertos productos, como los vinos, indicar el origen geográfico ha sido una práctica establecida durante muchos años, para otros este tipo de comercialización es más reciente [7].

En resumen, los consumidores demandan productos agrícolas o alimenticios con orígenes identificables y características específicas. En particular, los relacionados con el origen geográfico de los alimentos y sus condiciones de producción son motivo de gran preocupación [2]. En consecuencia a todo lo expuesto es que nace la Indicación Geográfica (IG), definida a nivel internacional en el Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio (ADPIC), siendo un sello o una designación que hace referencia al origen y permite valorar esta calidad específica al consumidor. Se supone que los productos con el sello de IG cumplen con las características vinculadas al origen, definidas colectivamente por los productores, las que a menudo son prácticas tradicionales que contribuyen a la conservación de la diversidad de los alimentos, a la valorización de los recursos locales, al mantenimiento de las tradiciones, al refuerzo de la organización local, a la lucha contra el éxodo rural y a ofrecer una gama más amplia de productos a los consumidores.

1.2.2. Definiciones y tipos de indicaciones geográficas: PDO, PGI y TSG

La Organización Mundial de la Propiedad Intelectual (OMPI) define una indicación geográfica (IG) como "un signo utilizado en productos que tienen un origen geográfico específico y poseen cualidades, una reputación o características que son esencialmente atribuibles a ese origen" [8]. La indicación geográfica consiste en el nombre del lugar de origen de los productos que poseen cualidades derivadas de su lugar de producción y en los que influyen factores geográficos específicos, como el clima y el suelo. El hecho de que un signo sea reconocido como indicación geográfica depende de la legislación nacional de que se trate. Las indicaciones geográficas pueden ser utilizadas para un gran número de productos, ya sean naturales, agrícolas o de fabricación.

La denominación de origen es un tipo especial de indicación geográfica que por lo general consiste en un nombre geográfico o una designación tradicional utilizada para productos que poseen cualidades o características específicas que cabe atribuir principalmente al entorno geográfico de producción. En el concepto de indicaciones geográficas quedan comprendidas las denominaciones de origen.

La protección de las indicaciones geográficas se contempla en tratados internacionales y legislaciones nacionales, sin la debida autorización no pueden utilizarse indicaciones geográficas respecto de productos que no sean originarios del lugar designado por dichas indicaciones. Las acciones aplicables en casos de infracción abarcan desde mandamientos judiciales hasta el pago de indemnizaciones por daños y perjuicios, multas y, en casos graves, penas de prisión.

En el plano internacional las indicaciones de origen están protegidas. En varios tratados administrados por la OMPI se contempla la protección de las indicaciones geográficas, en particular, en el Convenio de París para la Protección de la Propiedad Industrial, de 1883, y el Arreglo de Lisboa relativo a la Protección de las Denominaciones de Origen y su Registro Internacional. A esos tratados vienen a añadirse los artículos 22 a 24 del Acuerdo sobre los

Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio (Acuerdo sobre los ADPIC) en los que se contempla la protección de internacional de las indicaciones geográficas en el marco de la Organización Mundial del Comercio (OMC) [9].

Según la OMPI, los derechos de propiedad intelectual se rigen por el principio de territorialidad. Los efectos de un derecho obtenido en una jurisdicción particular se limitan al territorio de esa jurisdicción. Por lo tanto, cuando se obtiene un derecho sobre una IG en una jurisdicción, se protege allí pero no en el extranjero. Existen tres formas principales de proteger una IG en el extranjero: (1) obteniendo protección directamente en la jurisdicción correspondiente, (2) aprovechando los acuerdos bilaterales/plurilaterales celebrados entre países y (3) a través del Sistema de Lisboa de la OMPI para el Registro Internacional de denominaciones de origen.

A finales del siglo XX, la Unión Europea (UE) reconoció y apoyó el potencial de diferenciar productos de calidad a nivel regional. La legislación relativa a los sistemas de calidad se desarrolló paso a paso, desde el desarrollo de disposiciones especiales para vinos de calidad en la década de 1980 en adelante. En 1992, se adoptó la primera legislación europea sobre indicaciones geográficas y denominación de origen protegida para productos agrícolas y alimenticios (excluidas las bebidas espirituosas y los productos vitivinícolas), inspirada en los sistemas nacionales existentes, por ejemplo, el AOC francés (Appellation d'Origine Contrôlée) y el italiano DOC (Denominación de Origen Controlada) [10,11]. El primer reglamento sobre indicaciones geográficas se adoptó en la UE en 1992 para armonizar los diversos instrumentos de protección existentes en algunos Estados miembros y crear un sistema de registro y protección de nombres compatible con el mercado único común [12]. Desde entonces, la legislación de la UE establece requisitos estrictos que garantizan los estándares de todos los productos europeos. La UE creó tres etiquetas: la Denominación de Origen Protegida (Protected Designation of Origin, PDO), la Indicación Geográfica Protegida (Protected Geographical Indication, PGI) y la Garantía de Especialidad Tradicional (Traditional Specialty

Guarantee, TSG) que debe seguir estructuras de calidad para identificar productos y alimentos cultivados y producidos bajo especificaciones exactas. Estos sistemas fomentan la producción agrícola diversa, protegen los nombres de los productos contra el mal uso y la imitación, y ayudan a los consumidores brindándoles información sobre el carácter específico de los productos etiquetados [2]. Su adopción reserva el uso del nombre exclusivamente a los operadores que cumplan con todos los requisitos de las especificaciones.

La denominación de origen protegida o controlada DOC (PDO su sigla en inglés), se utiliza para diferenciar el producto por su lugar de origen. Esta herramienta cubre los productos agrícolas y alimenticios que se producen, procesan y preparan en un área geográfica dada (un lugar específico, región, o en casos excepcionales, un país) utilizando conocimientos reconocidos. Esto significa que se considera que el producto ha adquirido sus rasgos fundamentales esencialmente o exclusivamente debido a su lugar de origen. La naturaleza y el factor humano también se consideran elementos importantes; por lo tanto, toda la transformación de un producto crudo a un producto terminado debe llevarse a cabo en la misma área geográfica [2,13]. En lo que se refiere a los vinos y a las bebidas espirituosas de origen vitivinícola, la denominación de origen reconocida: designa un producto cuya tipicidad y características se deben exclusiva o esencialmente al medio geográfico que comprende los factores naturales y los factores humanos, y está subordinada a la cosecha así como a la transformación en el país, la región, el lugar o el área definida [14].

La indicación geográfica protegida (PGI su sigla en inglés) protege los productos agrícolas y alimenticios estrechamente vinculados al área geográfica (un lugar, región o país específico), cuya calidad, reputación u otras características dadas son esencialmente atribuibles a su origen geográfico, y al menos una de las etapas de producción, procesamiento o preparación se lleva a cabo en el área geográfica definida mientras que las materias primas utilizadas en la producción pueden provenir de otra región [2,10]. En lo que se refiere a los vinos, el reconocimiento de este nombre, está relacionado con una calidad y/o con una

característica del producto atribuidas al medio geográfico que comprende los factores naturales o los factores humanos, y está subordinada a la cosecha de la uva en el país, la región, el lugar o el área definida [14].

Por último, la garantía de especialidad tradicional (TSG su sigla en inglés) indica el carácter tradicional de los alimentos, ya sea en la composición o en los medios de producción que los diferencian de otros productos de la misma categoría. Describe un producto alimenticio específico que resulta de un modo de producción, procesamiento o composición correspondiente a la práctica tradicional para ese producto alimenticio o que se produce a partir de materias primas o ingredientes que son los utilizados tradicionalmente [2,13].

La incorporación de estos sistemas de indicación geográfica en alimentos no es sólo un asunto europeo. Actualmente los productores de países fuera de la UE pueden unirse a los sistemas de certificación de calidad de acuerdo con las normas de la Organización Mundial del Comercio sobre comercio internacional [11]. En muchos casos, el sistema de PDO/PGI de la UE funciona en paralelo con el sistema utilizado en cada país, y en algunos casos está subordinado al sistema de denominación que ya se instituyó. México fue el primer país no europeo en establecer un sistema de protección IG en 1974. Brasil y Perú aprobaron leyes sobre IG en 1996, seguidos de Corea del Sur e India en 1999, Colombia en 2000 y Chile en 2005 [2].

Finalmente, la incorporación de estos sistemas se realiza gráficamente mediante signos o sellos en los productos de distintos mercados, que certifican los atributos que ostentan a merced a su vinculación con su origen y simultáneamente con una protección y diferenciación que les confiere mayor valor agregado. A continuación se presenta en la Figura 1.8 los logos de los sistemas de calidad de origen para los productos originarios de Argentina.



Figura 1.8. Logos de DO e IG en el etiquetado de productos de Argentina.

1.2.3. Enfoque vitivinícola: concepto de terruño o terroir

Las primeras indicaciones geográficas (IG) del mundo fueron en el sector vitivinícola y se centraron en la delimitación de la ubicación de la producción. Históricamente, los vinos que se comercializaron fueron los vinos de mejor calidad que subsistieron a viajes con extensas distancias sin perder características sustanciales, como sabor, y cuyos precios más altos cubrieron los costos de transporte. Por lo tanto, el comercio y la calidad estaban vinculados. Garantizar la calidad era importante para el comercio y un instrumento importante que se utilizó fue el vínculo de la calidad del producto con la región donde se produjo, naciendo el concepto de “terruño” o “terroir” asociado al origen del vino [15]. El papel del terruño ha sido importante históricamente en la regulación del vino, especialmente en Europa con ejemplos como los vinos de Borgoña en el siglo XV, los vinos de Oporto y los vinos de Chianti en el siglo XVIII, y el vino de Champagne a principios del siglo XX reconocidos mundialmente por su origen.

El término francés "terroir" (de la palabra *terre*, que significa suelo) se define como la interacción entre el entorno físico y biológico con las vides, en el que la geología, la geomorfología, el suelo, las condiciones climáticas, la biología de la vid, y la actividad humana proporcionan características distintivas y de alta calidad para los productos originados en un área específica [16–18].

De acuerdo a la OIV, “El Terroir vitivinícola es un concepto que se refiere a un espacio sobre el cual se desarrolla un saber colectivo de las interacciones entre un medio físico y biológico identificable y las prácticas vitivinícolas aplicadas, que confieren unas características

distintivas a los productos originarios de este espacio”. “El Terroir incluye características específicas del suelo, de la topografía, del clima, del paisaje y de la biodiversidad” [19].

Para comprender cómo funciona el concepto de terroir o terruño, es esencial tener en cuenta las interacciones entre los factores que contribuyen en él. Los factores del entorno natural deben considerarse en términos de su interacción con la vid. Los factores humanos, como historia, socio-economía, así como las técnicas vitícolas y enológicas, también forman parte del terruño.

Factores Humanos

Es importante tener en cuenta el factor humano en el terroir porque no existe un viñedo sin la intervención del hombre. Estos factores tienen una relación con los cuidados culturales en general y, por eso, dependen del viticultor. Ellos también tienen influencia en la selección de las variedades de la vid, el rendimiento en uva del cepaje, poda, riego, abonaduras, entre otros, interviniendo entonces en la calidad del producto obtenido [20]. Se necesitan conocimientos técnicos y experiencia para hacer vinos de alta calidad que expresen terruño.

Factores del entorno natural

Clima

El clima se considera un factor fundamental en la expresión del terroir. Si bien los vinos de alta calidad se cultivan en varios climas, es imposible definir el clima ideal para los vinos en términos de temperatura. El clima influye en la fisiología de la vid a través de las temperaturas, precipitaciones, déficit de presión de vapor, evapotranspiración, horas de sol y viento.

La vid es una planta perenne adaptada a una amplia gama de condiciones climáticas. El hábitat natural es el clima templado, pero se adapta también a climas cálidos y relativamente fríos [16].

Suelo

El suelo se considera uno de los componentes más importantes del terroir, y existe evidencia empírica que relaciona el vino con las condiciones específicas del suelo del viñedo. Las vides se pueden cultivar en una gran variedad de suelos. En suelos profundos y ricos, las vides son vigorosas y altamente productivas, pero generalmente se producen mejores vinos cuando éstas se cultivan en suelos pobres. El efecto del suelo sobre el comportamiento de la vid y la composición de la uva es complejo, su papel no puede explicarse sin desglosarlo en varios efectos secundarios: textura, composición mineral, suministro de agua a las vides, temperatura en la zona de la raíz y muchos otros factores, por lo que su estudio abarca desde una perspectiva geológica, pedológica y agronómica [20]. Explorando brevemente estos enfoques, la geología se ocupa de los tipos de rocas y la datación de estratos sedimentarios. La geología actúa indirectamente sobre la topografía. El tipo de suelo también está relacionado con el tipo de roca sobre la que se ha desarrollado. Los tipos de suelo se pueden mapear, en base a una clasificación pedológica. Algunos tipos de suelo, como los suelos de piedra caliza, son conocidos por producir vinos en general de alta calidad, mientras que otros, como los suelos sujetos a anegamiento, son conocidos por ser menos adecuados para producir vinos de calidad. Sin embargo, en todo el mundo, los vinos sobresalientes se cultivan en una gran variedad de suelos.

Para comprender el efecto del suelo en la viticultura, además es necesario tener en cuenta la interacción entre el suelo y la vid (enfoque agronómico). El suelo influye en el comportamiento de la vid y la calidad del vino a través de la temperatura en la zona de la raíz, y a través de los suministros de minerales y agua. Entre los minerales que el suelo suministra a la vid, el nitrógeno es obviamente el que más influye en el vigor, el rendimiento y la maduración de la uva. El efecto de otros minerales en el desarrollo de la vid y el potencial de calidad de la uva es mucho menos obvio, siempre que ni el exceso ni la deficiencia severa alteren la fisiología de la vid. Según algunos autores de la literatura, la composición mineral del suelo no parece ser, en general, un factor decisivo del terruño, tres elementos que la vid

recoge del suelo necesitan más atención: nitrógeno, potasio y calcio. El nitrógeno y el potasio influyen en el desarrollo de la vid y la composición de la uva. El calcio es un factor importante para mejorar la estructura del suelo.

Las condiciones de absorción de agua de la vid son un factor agronómico clave para comprender el efecto del terroir en el potencial de calidad de la uva, porque los principales factores del terroir están involucrados e interactúan (clima, suelo, vid). El suelo es el soporte de las raíces, la reserva de donde toman el agua y los nutrientes, y el medio en el que viven, de manera que sus características van a condicionar la reserva hídrica y la absorción de los nutrientes a lo largo del ciclo vegetativo [16].

Vinculadas al concepto de terroir en la vitivinicultura, las indicaciones geográficas aparecen como señales de calidad adicionales para el vino. Históricamente en Europa, y particularmente en Francia, el vino se vendía por su origen y las regiones vitivinícolas se han demarcado desde hace mucho tiempo en el sistema de Denominación de Origen Controlada o Protegida (o AOC en francés, Appellation d'Origine Contrôlée). Sin embargo, esta denominación se creó inicialmente sólo para proteger el nombre de una comuna o región vitivinícola, y su delimitación se basó en límites administrativos en lugar de estudios extensos del entorno natural. Progresivamente, se incluyeron algunas variables ambientales en las prácticas de delimitación (geología, tipo de suelo, susceptibilidad de los suelos al agua), y además los factores humanos. Hoy en día, básicamente, en el caso de los vinos, estas identificaciones geográficas garantizan que el producto está realizado bajo las normas que regulan dicha DOC., con uvas procedentes de esa zona o región, respetando los rendimientos y la forma de plantación de las vides, junto con los métodos enológicos estipulados. Cada DOC cuenta con sus organismos controladores que aseguran y avalan el cumplimiento de las normas que la reglamentan, obteniendo así un valor agregado.

1.2.4. Reglamentación y marco legal

A nivel internacional, diferentes organizaciones se han ocupado de clarificar y llegar a un consenso en las definiciones y normativas que han permitido una adecuada designación del origen del vino y su comercialización. Hasta hace relativamente escaso tiempo, la UE clasificaba la calidad del vino en dos categorías: vino de calidad producido en una región específica y vino de mesa. Estos fueron reemplazados en 2011 por las denominaciones PDO (Denominación de Origen Protegida) y PGI (Indicación Geográfica Protegida) explicadas anteriormente. La Organización Internacional de la Viña y el Vino (OIV) es quién tiene la tarea de garantizar la autenticidad de los productos derivados de la vid, sobre todo cara a los consumidores y en particular en cuanto a las menciones de etiquetado, la protección de las indicaciones geográficas y en particular de las áreas vitivinícolas y las denominaciones de origen, designadas por nombres geográficos o no, que les son asociados, en la medida en que no cuestionan los acuerdos internacionales en materia de comercio y propiedad intelectual, la mejora de los criterios científicos y técnicos de reconocimiento y protección de las obtenciones vegetales vitivinícolas.

En lo que respecta a Argentina, para los vinos la protección de las indicaciones geográficas cuenta con una historia más reciente. Los intentos iniciales se remontan a principios de la década de 1990, cuando se estableció el primer esquema de calidad del vino relacionado con el lugar de producción. La versión final se estableció en la Ley Nº 25.163 aprobada en 1999. El objeto de la ley Nº 25.163 es establecer un sistema de reconocimiento, protección y registro de nombres geográficos argentinos para designar el origen de los vinos y de las bebidas espirituosas de naturaleza vínica. Dicha Ley establece tres categorías para la designación y presentación de vinos y bebidas espirituosa: la Indicación de Procedencia (IP) para los vinos de mesa o vinos regionales y la Indicación Geográfica (IG) y Denominación de Origen Controlada (DOC) para vinos de calidad superior [21,22]. La Autoridad de Aplicación de esta ley es la Secretaría de Agricultura, Ganadería, Pesca y Alimentación del Ministerio

Economía y Obras y Servicios Públicos, a través del Instituto Nacional de Vitivinicultura (INV), con jurisdicción en todo el territorio la Nación Argentina.

Actualmente en Argentina existen sólo dos DOC en vinos encontrados en la provincia de Mendoza:

- DOC Lujan de Cuyo, desde el año 2005
- DOC San Rafael, desde el año 2007

En el caso de otros productos alimentarios, existe en Argentina el Régimen Leal de las Indicaciones Geográficas y Denominaciones de Origen de Productos Agrícolas y Alimentarios constituido por la Ley Nº25380 y su modificatoria Ley Nº 25.966. A su vez, la Secretaría de Agregado de Valor, a través del Registro Nacional de IG y DO (creado por Resolución Nº 587/2010) es el organismo del Estado Nacional donde se tramitan las solicitudes de reconocimiento y registro de estas herramientas de diferenciación y agregado de valor para los productos con cualidades diferenciales debidas al lugar de producción o elaboración. Estas normas fueron dictadas considerando los estándares fijados en el Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio (ADPIC) con el objetivo de proteger los productos argentinos en el mercado global y con la aspiración de posicionarlos y valorizarlos.

Con respecto a otros productos distintos del vino, hoy en día existe PDO en productos alimenticios y bebidas. A pesar de esto, los jugos de frutas no se caracterizan específicamente por esta importante etiqueta de calidad y autenticidad. No existen jugos etiquetados con protección de indicación geográfica. El mercado de los zumos de frutas se centra en los tipos de sabores únicos, principalmente de naranja y manzana, así como en los zumos multifrutas elaborados con una mezcla de sabores. Incluso con los jugos hechos de una sola fruta, la industria de los jugos argumenta que las fluctuaciones de cosecha y las diferencias en la calidad y disponibilidad de frutas requieren una mezcla de jugos de diferentes orígenes para garantizar una calidad constante del consumidor. Los productores de jugos consideran que no

es viable económicamente elaborar jugos de frutas con la denominación de origen porque la materia prima proviene de diferentes partes del mundo. Sin embargo hay un mercado creciente para jugos de valor agregado basados en las propiedades nutritivas y saludables inherentes de las frutas y verduras [23].

1.2.5. Referencias

- [1] International Organization for Standardization, (ISO), ISO9000:2000 and ISO9000:2005, Quality management systems., (n.d.). www.iso.org.
- [2] S. Armenta, M. de la Guardia, Analytical Approaches for the Evaluation of Food Protected Designation of Origin, in: Adv. Food Traceability Tech. Technol., Elsevier, 2016: pp. 275–301. doi:10.1016/B978-0-08-100310-7.00015-6.
- [3] A. Gonzalez, S. Armenta, M. de la Guardia, Trace-element composition and stable-isotope ratio for discrimination of foods with Protected Designation of Origin, TrAC - Trends Anal. Chem. 28 (2009) 1295–1311. doi:10.1016/j.trac.2009.08.001.
- [4] G.P. Danezis, A.S. Tsagkaris, F. Camin, V. Brusic, C.A. Georgiou, Food authentication: Techniques, trends & emerging approaches, TrAC - Trends Anal. Chem. 85 (2016) 123–132. doi:10.1016/j.trac.2016.02.026.
- [5] M.M. Aung, Y.S. Chang, Traceability in a food supply chain: Safety and quality perspectives, Food Control. 39 (2014) 172–184. doi:10.1016/j.foodcont.2013.11.007.
- [6] M. Woolfe, T. Ditton, Food origin labelling legislation and standards, in: New Anal. Approaches Verif. Orig. Food, Elsevier, 2013: pp. 12–37. doi:10.1533/9780857097590.1.12.
- [7] A. Roßmann, 3 – New approaches for verifying the geographical origin of foods, in: New Anal. Approaches Verif. Orig. Food, 2013: pp. 41–59. doi:10.1533/9780857097590.2.41.
- [8] WIPO, World Intellectual Property Organization. Switzerland, Geneva, (2015).
- [9] OMPI, Organización Mundial de la Propiedad Intelectual, Acerca de las indicaciones geográficas, (2020). https://www.wipo.int/geo_indications/es/about.html (accessed

January 14, 2020).

- [10] A. Hajdukiewicz, European Union agri-food quality schemes for the protection and promotion of geographical indications and traditional specialities: an economic perspective, *Folia Hort.* 26 (2014) 3–17. doi:10.2478/fhort-2014-0001.
- [11] C. Dias, L. Mendes, Protected Designation of Origin (PDO), Protected Geographical Indication (PGI) and Traditional Speciality Guaranteed (TSG): A bibliometric analysis, *Food Res. Int.* 103 (2018) 492–508. doi:10.1016/j.foodres.2017.09.059.
- [12] EU, Council, Council Regulation (EEC) No 2081/92 “Protection of geographical indications and designations of origin for agricultural products and foodstuffs,” *Off. J. Eur. Union.* 208 (1992) 0001–0008.
- [13] A. Krystallis, P. Chrysochou, T. Perrea, N. Tzagarakis, A Retrospective View on Designation of Origin Labeled Foods in Europe, *J. Int. Food Agribus. Mark.* 29 (2017) 217–233. doi:10.1080/08974438.2017.1307156.
- [14] International Organization of Vine and Wine (OIV), Norma Internacional para el etiquetado de los vinos, (2015). <http://www.oiv.int/es/la-organizacion-internacional-de-la-vina-y-el-vino/acerca-de-la-oiv>.
- [15] G. Meloni, J. Swinnen, Trade and terroir. The political economy of the world’s first geographical indications, *Food Policy.* 81 (2018) 1–20. doi:10.1016/j.foodpol.2018.10.003.
- [16] C. van Leeuwen, 9 – Terroir: the effect of the physical environment on vine growth, grape ripening and wine sensory attributes, in: *Manag. Wine Qual.*, 2010: pp. 273–315. doi:10.1533/9781845699284.3.273.
- [17] F. Foroni, M. Vignando, M. Aiello, V. Parma, M.G. Paoletti, A. Squartini, R.I. Rumiati, The smell of terroir! Olfactory discrimination between wines of different grape variety and different terroir, *Food Qual. Prefer.* 58 (2017) 18–23. doi:10.1016/j.foodqual.2016.12.012.

- [18] C. Zucca, R. Biancalani, S. Kapur, E. Akça, P. Zdruli, L. Montanarella, F. Nachtergaele, *Soil Security for Ecosystem Management*, 8 (2014) 31–59. doi:10.1007/978-3-319-00699-4.
- [19] The General Director of the OIV, G. Assembly, Resolution OIV/VITI 333/2010 Definition of vitivinicultural “Terroir,” (2010) 1.
- [20] C. Van Leeuwen, G. Seguin, The concept of terroir in viticulture, *J. Wine Res.* 17 (2006) 1–10. doi:10.1080/09571260600633135.
- [21] E. Defrancesco, J. Estrella, A. Gennari, Would ‘New World’ wines benefit from protected geographical indications in international markets ? The case of Argentinean Malbec, *Wine Econ. Policy.* 1 (2012) 63–72. doi:10.1016/j.wep.2012.08.001.
- [22] Instituto Nacional de Vitivinicultura (INV), Ley de Denominación en Origen N° 25.163. Decreto reglamentario N°57/2004, Argentina, (2004).
- [23] M. Díaz-García, M.R. Castelar, J.M. Obón, Juices and Non-Alcoholic Beverages, in: *Compr. Anal. Chem.*, Elsevier, 2013: pp. 439–459. doi:https://doi.org/10.1016/B978-0-444-59562-1.00017-7.

1.3. ANÁLISIS MULTIELEMENTAL PARA LA DESIGNACIÓN DE ORIGEN

1.3.1. El perfil elemental como 'fingerprint' de los alimentos

Como ya se expuso anteriormente, la determinación de la autenticidad de los alimentos es un tema de relevancia para el control de calidad y seguridad de los mismos. La autenticidad es un procedimiento que tiene como objetivo resaltar las diferencias entre los alimentos del mismo tipo que se difieren en términos del origen (especie, geográfico o genético), método de producción (por ejemplo, convencional, orgánico, procedimientos tradicionales) o tecnologías de procesamiento (por ejemplo, irradiación, congelación, calentamiento por microondas). La declaración de atributos de calidad específicos en productos de alto valor es de particular interés ya que estos productos a menudo son objeto de etiquetado fraudulento. La identificación de la procedencia es un tema importante para la inocuidad y calidad de los alimentos y la protección del consumidor, así como el cumplimiento de la legislación nacional y las normas internacionales. La autenticidad depende de las características medibles que puedan discriminar los alimentos de diferentes orígenes. Para dicho cometido, se pueden emplear parámetros inorgánicos como así también compuestos orgánicos o características espectrales. Sin embargo, gracias a su papel conocido en los procesos geoquímicos, los elementos inorgánicos son especialmente adecuados como marcadores para la autenticación, siendo particularmente ventajoso determinar el mayor número posible de elementos (perfil multielemental o huellas dactilares) para tener el mayor número posible de variables para la discriminación [1–3]. A su vez, el rango normal de compuestos orgánicos en los alimentos varía con algunos parámetros agrícolas, como la fertilización, las condiciones climáticas en el año de cultivo, la historia de los campos y la variedad o especies, así como la ubicación geográfica y las características del suelo, por lo que a veces es difícil ser definitivo sobre la autenticidad de un material a partir de la determinación de componentes orgánicos [4].

El enfoque de huella dactilar ('fingerprint' en inglés) de alimentos, generalmente se basa en una selección de muestras con el propósito de diferenciar o clasificarlas. La huella dactilar de alimentos posee dos enfoques de análisis: específico (targeted) y no específico (non-targeted), con el objetivo intrínseco de detectar y reconocer componentes de la matriz que representan patrones para la diferenciación o clasificación de las muestras [5]. Entonces se puede definir el concepto como un marcador o conjuntos de marcadores moleculares que representan un estado o condición alimentaria característica, que permite una discriminación más efectiva del producto [6].

Dentro del enfoque de análisis específico, la determinación del contenido mineral de alimentos surge como una herramienta de gran importancia a la hora de asegurar la autenticidad de alimentos en base a su procedencia. El contenido de minerales refleja el tipo de suelo de un área de cultivo y las condiciones ambientales de crecimiento para la producción de alimentos [7].

Casi todo comienza desde el suelo. Es posible demostrar que la mayoría de los alimentos, ya sean de origen vegetal (frutas, verduras, aceites, etc.) o de origen animal (carnes, leche, productos lácteos, etc.), derivan su contenido elemental en primer lugar del suelo. En las concentraciones de elementos traza o en las relaciones de isótopos, el papel de cada elemento químico en la interacción entre el suelo y las plantas es la clave para comprender si un determinado elemento podría actuar como un marcador para la clasificación. En este contexto, los elementos se pueden dividir en tres grupos [1]:

- Macro-esenciales o nutrientes: estos elementos son indispensables para la vida de las plantas y, por lo tanto, se asumen activamente desde el suelo. Los elementos alcalinos y alcalinos térreos y algunos metales de transición pertenecen a este grupo.
- Micro-esenciales: estos elementos tienen un papel menor en la fisiología de las plantas, por lo que generalmente se asumen de manera pasiva. Los elementos de tierras raras (REE) pertenecen a este grupo.

- Tóxicos: estos elementos son perjudiciales para las plantas que disponen en algunos casos de rutas metabólicas para evitar recogerlos del suelo o eliminarlos rápidamente. Los metales pesados como cadmio, mercurio, talio o plomo pertenecen a este grupo.

El perfil elemental está enfocado entonces en grupos de elementos que incluyen macro-elementos (Ej. calcio, sodio, potasio), oligoelementos (Ej. selenio, zinc, cobre), elementos de tierras raras (REE: cerio, samario y lantano) u otros elementos ultra-traza (Ej. metales preciosos como platino u oro). Los REE y los metales preciosos tienen un gran potencial para la determinación del origen geográfico porque reflejan las condiciones del medio ambiente local, sus huellas dactilares están directamente vinculadas a la geología del área y podrían ser influenciadas mínimamente por las diferentes prácticas agrícolas y el año de cosecha [2]. En particular, los elementos de tierras raras se consideran herramientas útiles como marcadores de trazabilidad. Los geoquímicos los usan como marcadores geológicos en la identificación del origen de las rocas, a su vez los REE muestran características similares en los sistemas biológicos, ya que no parecen tener un papel activo en la fisiología de las plantas. Estudios apuntan a que las plantas tienden a absorber iones REE del suelo con poco o ningún fraccionamiento de la distribución original, y este comportamiento podría transmitirse a los alimentos con cadenas de producción simples [1].

Las plantas captan los elementos químicos del suelo. La fertilización, cosecha, origen botánico, tipo de suelo, contaminación y año de producción causan variaciones de las concentraciones elementales. Sin embargo, estas variaciones son menores que la variación observada entre las áreas de producción y las regiones geográficas. La composición elemental de los alimentos de origen animal refleja, en cierta medida, el contenido mineral del forraje y la vegetación que los animales comen [2].

En los esquemas de trazabilidad, los marcadores ideales son aquellos elementos cuyas concentraciones no están influenciadas por fuentes secundarias. Las concentraciones de tales elementos en los alimentos o en sus materias primas pueden reflejar proporcionalmente las

concentraciones en el suelo, debido al hecho de que se asumen sin la intervención activa de las plantas. Cuando se asume un determinado elemento de manera preferencial porque es un nutriente, o se excluye de la absorción porque es tóxico, tenemos un fraccionamiento de la distribución original en el suelo [1].

Un determinado elemento podría actuar como marcador de autenticidad incluso si su concentración se originara desde diversas fuentes, siempre que las contribuciones externas a la natural sean sistemáticas y representativas de características geográficas, varietales o tecnológicas. En este caso, es difícil establecer cuáles son los mejores marcadores, es decir, los elementos más adecuados para cada esquema de clasificación. Los elementos adecuados para la clasificación de un determinado alimento podrían resultar insuficientes para la clasificación de otro alimento. Los elementos mayoritarios, minoritarios, traza o ultra-traza pueden ser marcadores efectivos al mismo tiempo, pero dado que los elementos mayoritarios/minoritarios pueden ser componentes activos en los procesos tecnológicos involucrados en las cadenas de producción, es decir, en cada grupo de elementos (es decir, metales alcalinos, alcalinotérreos o de transición), los componentes más abundantes son generalmente menos confiables que los más escasos como marcadores de autenticación [1].

En resumen, los elementos mayoritarios y minoritarios parecen funcionar mejor en los esquemas de autenticación basados en la discriminación varietal o tecnológica, mientras que los elementos traza y ultra-traza son en su mayoría adecuados para discriminar sobre una base geográfica, lo que es consistente con su papel como marcadores geoquímicos como tal.

La gran diversidad de regiones en donde se desarrolla la actividad vitivinícola genera la necesidad de establecer la procedencia de los productos elaborados. Así, el vino se identifica con su territorio de origen. Para ello, se han utilizado diferentes parámetros fisicoquímicos a los efectos de diferenciar los vinos según su origen, sin embargo, los metales son los más utilizados para demostrar la relación entre el vino, el jugo de uva y la composición del suelo. De este modo, el enfoque basado en el perfil mineral de los vinos es uno de los métodos de

'fingerprint' específicos (target). En la Figura 1.9 se pueden observar los métodos específicos y no específicos de análisis para monitorear y evaluar la denominación de origen del vino [8].

El vino, desde un punto de vista químico, es una matriz compleja con agua y etanol como componentes principales, que contiene Fe, K, Mg, Ca, Na y Al a nivel de ppm y alrededor de 15 elementos que están comúnmente presentes a nivel de ppb, por lo que su contenido en metales es diversa y de diferente fuente de procedencia.

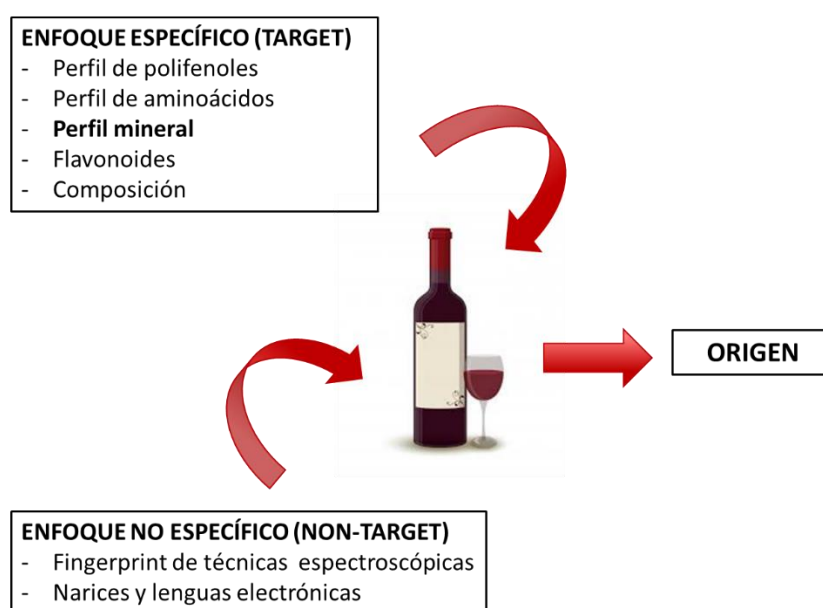


Figura 1.9. Enfoques para el monitoreo y evaluación de la designación de origen en vinos.

Las uvas reciben los elementos traza captados por la planta de vid desde suelo, por lo que su composición elemental se correlaciona con las características geológicas de las regiones de producción. Los elementos minerales no se metabolizan ni transforman durante los procesos enológicos y permanecen casi inalterados de los mostos a los vinos [9]. Por lo tanto, esta composición mineral refleja el ambiente en el que se cultivan las vides, y como consecuencia, los productos de la vid (es decir, uvas, jugo, vino) están influenciados por la composición del suelo [10].

Hay dos fuentes principales que contribuyen a la presencia de diferentes metales en el vino. Los metales de origen primario y natural provienen del suelo donde se cultivan las vides y llegan al vino a través de las uvas suministrando la mayor parte de los iones presentes en el

vino. La concentración de estos metales está relacionada con la madurez de las uvas, su variedad, el tipo de suelo en el viñedo y las condiciones climáticas durante su crecimiento. En segundo lugar, tenemos las fuentes derivadas de la actividad humana, como prácticas de cultivo (fertilización, tratamientos fitosanitarios, etc.), prácticas enológicas (maquinaria, tuberías, uso de agentes de clarificación, aditivos, etc.), contaminación ambiental (automóviles, fábricas, etc.) y prácticas fraudulentas (uso de productos químicos prohibidos) [10,11]. En la Figura 1.10 se presentan esquemáticamente las dos fuentes principales de metales en el vino.

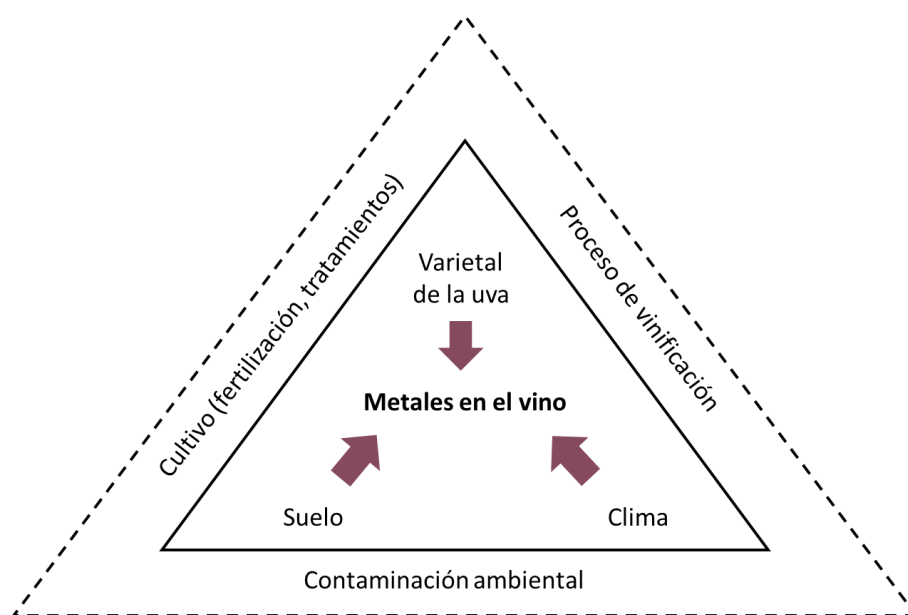


Figura 1.10. Fuentes endógenas o primarias (línea continua) y exógenas o secundarias (línea punteada) de metales en el vino [10].

La vid absorbe las sustancias minerales del suelo de acuerdo con sus necesidades fisiológicas, lo que implica un fraccionamiento de la distribución: se favorece la de elementos esenciales (es decir, Ca, Na, K, etc.), se reduce la de elementos tóxicos (es decir, Cd, Hg, Pb, etc.), y los elementos no esenciales (es decir, Ba, Rb, Sr, etc.) pueden absorberse debido a su afinidad química con los elementos esenciales [1].

Los metales juegan un papel importante que influye en la calidad del vino, ya que afectan las características organolépticas del vino, incluidos el sabor, la frescura, el aroma, el color y el sabor, principalmente debido a la formación de precipitados (levadura, clarificación y

filtración sedimentos) o al enturbiamiento durante la fermentación, maduración y almacenamiento del vino [10,11].

La concentración de macro y micronutrientes (Ej. K, Ca, Fe, Cu y Zn) varía ampliamente durante el procesamiento de las uvas hasta transformarse en vino. Por el contrario, los metales alcalinotérreos como el litio, rubidio y algunos lantánidos (es decir, elementos de tierras raras) han mostrado menos cambios con el procesamiento y el tiempo, y por lo tanto son más relevantes para autenticar el origen geográfico de los vinos [10]. En cuanto a la conversión de uvas en mosto, algunos estudios demostraron que la distribución de REE en el suelo se mantiene bien en el mosto correspondiente, lo cual está de acuerdo a lo expuesto anteriormente considerando estos elementos como marcadores de trazabilidad [1].

Los metales son entonces muy buenos indicadores del origen del vino y pueden usarse para garantizar su autenticidad, ya que no se metabolizan ni modifican durante la vinificación y reflejan la composición promedio de los suelos de los viñedos. Si bien existen factores como las prácticas agrícolas, los cambios climáticos, contaminación ambiental y los procesos de vinificación que podrían cambiar el patrón primario de metales en el vino y poner en peligro la relación entre el vino y la composición del suelo, el análisis del perfil mineral como huella dactilar o fingerprint del vino actúa como una especie de enfoque no dirigido tendiente a determinar el mayor número posible de elementos que combinados con herramientas quimiométricas ofrecen ventajas para caracterizar, clasificar y autenticar los vinos.

1.3.2. Técnicas analíticas elementales

En la actualidad existen numerosas herramientas analíticas disponibles para la determinación de elementos traza y ultra-traza. Las técnicas de elección para obtener la huella dactilar elemental de los alimentos son principalmente aquellas con capacidad de detección multielemental, como las basadas en la fuente ICP: espectroscopia de emisión óptica de plasma acoplado inductivamente (ICP-OES) y espectrometría de masas de plasma acoplado inductivamente (ICP-MS). Como técnicas analíticas alternativas para determinar el origen

geográfico de los alimentos, se puede utilizar la espectrometría de absorción atómica clásica (AAS) basada en sistemas de atomización electrotérmica (ETAAS) o de llama (FAAS). La espectrometría de emisión también se ha utilizado como técnica analítica para garantizar el origen geográfico de los alimentos, como la espectrometría de emisión atómica de llama (FAES). Otros enfoques menos comunes incluyen el análisis radioquímico de activación de neutrones (RNAA) y el análisis instrumental de activación de neutrones (INAA). En menor medida, la fluorescencia de rayos X (XRF) es otra técnica apropiada para la clasificación de alimentos con PDO basada en el perfil mineral de las muestras [3,7].

A continuación se explayará sobre la técnica de ICP-MS, la cual fue la herramienta analítica utilizada para llevar a cabo el análisis multielemental en la presente tesis doctoral.

1.3.2.1. Espectrometría de masas con plasma acoplado inductivamente

El acoplamiento de la espectrometría de masas a los métodos de atomización/ionización de muestras en plasma, proporciona una de las técnicas más sensibles para la determinación de elementos traza y ultra-traza: la espectrometría de masas con plasma acoplado inductivamente (ICP-MS, del inglés “Inductively coupled plasma mass spectrometry”). Esta técnica se ha utilizado desde la década de 1980 en varios campos de aplicación, y desde la década de 1990 ha demostrado su potencialidad en la autenticación de alimentos [1].

El ICP-MS es una técnica basada en la medición de las masas atómicas de un determinado isótopo de un elemento dado. Las mediciones se realizan en función de la relación masa/carga (m/z) obtenidas para cada isótopo. Esta técnica es una herramienta poderosa para la determinación cuantitativa de una gama de metales y no metales (elementos inorgánicos) en una amplia variedad de muestras en niveles de concentración de trazas (ppb – ppm) y ultra-trazas (ppq – ppb). Además ofrece otras ventajas entre las que encontramos la obtención de espectros sencillos, generalmente únicos y con frecuencia de fácil interpretación, las capacidad de medir relaciones isotópicas atómicas, velocidad de análisis, análisis de

muestras tanto sólidas como líquidas y su característica de gran utilidad para la realización de esta tesis doctoral, su capacidad multielemental de análisis [12,13].

En la Figura 1.11 se presentan los elementos posibles de ser determinados por ICP-MS y los rangos de límites de detección alcanzados para cada uno de ellos.

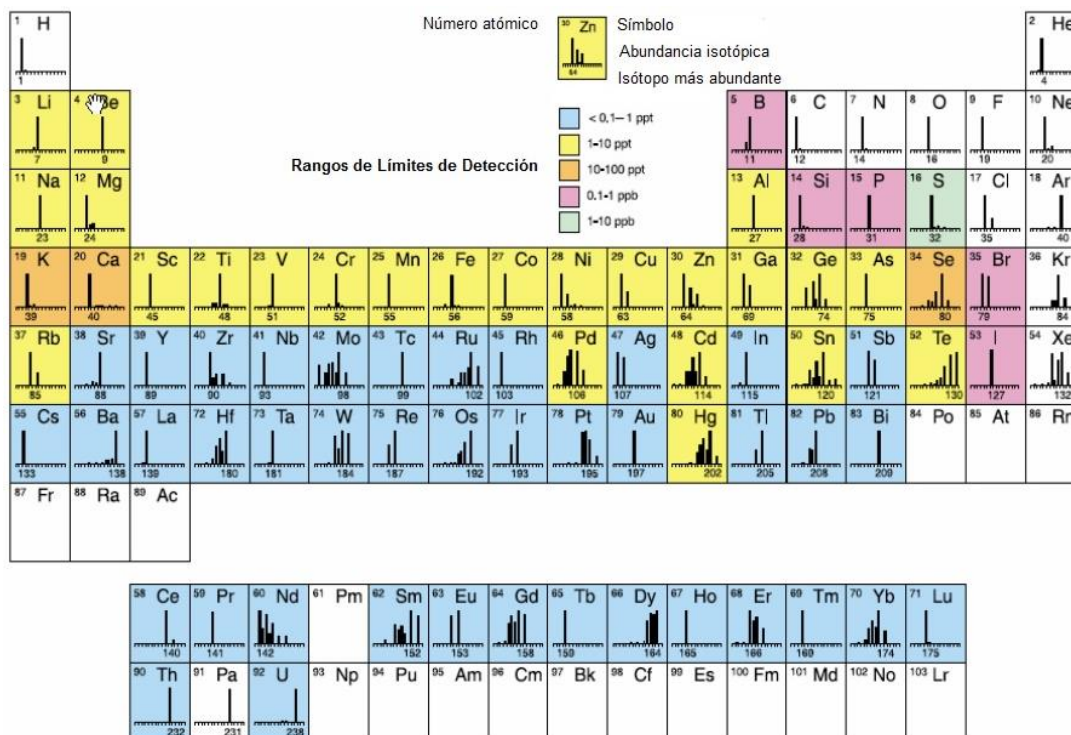


Figura 1.11. Elementos que pueden determinarse por ICP-MS [14].

Entre las limitaciones que presenta ICP-MS, se puede mencionar la influencia que tienen los efectos de matriz, la necesidad de mantener baja la concentración de sólidos disueltos para evitar problemas de obstrucción del nebulizador neumático, el elevado costo, la deriva del instrumento que puede ser del 5 al 10% por hora y la necesidad de un alto nivel de experiencia del personal. A pesar de dichas desventajas, el ICP-MS es una técnica de gran uso y ha encontrado una amplia aplicación para el análisis de una gran variedad de muestras, incluyendo agua y alimentos, así como muestras geoquímicas, ambientales y biológicas, entre otras [15].

1.3.2.1.1. Instrumentación y funcionamiento

Las partes fundamentales de un instrumento de ICP-MS son las siguientes: 1) sistema de introducción de muestras, 2) generación de iones (ICP), 3) interfase de acondicionamiento, 4) sistema de focalización de iones, 5) discriminador de iones o analizador de masas y 6) detección [16]. En la Figura 1.12 se muestra un esquema genérico de un sistema ICP-MS.

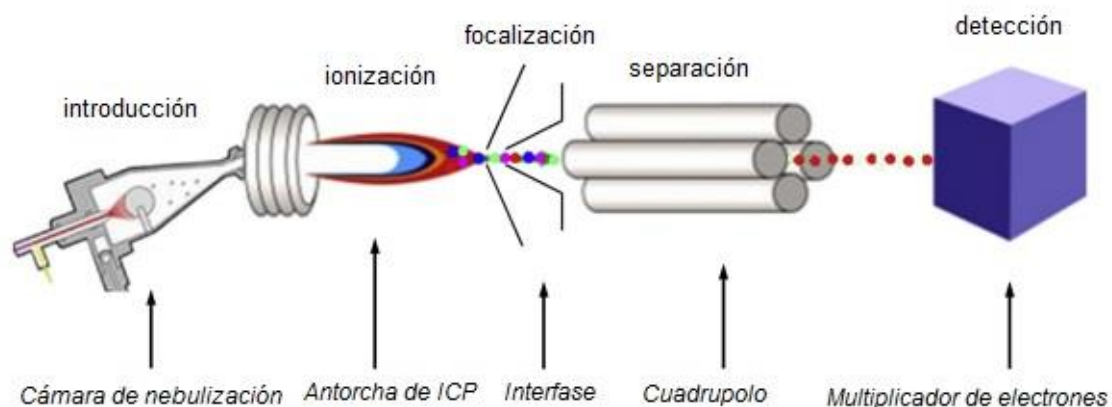


Figura 1.12. Esquema de un instrumento ICP-MS [1].

Introducción de la muestra

En la técnica de ICP-MS es posible la introducción continua de muestras líquidas, sólidas o gaseosas, sin embargo, los instrumentos están diseñados principalmente para analizar líquidos. Las muestras sólidas se pueden analizar directamente mediante vaporización electrotérmica o ablación con láser, y en cuanto a la introducción continua de muestras gaseosas al ICP-MS, puede recurrirse a la reacción de generación de hidruros, la cual resulta de gran utilidad para elementos como As, Bi, Se, Pb, Sb.

Las muestras líquidas se pulverizan primero con un nebulizador antes de ser introducidas en el ICP. Una configuración común utiliza un automuestreador y una bomba peristáltica para introducir la muestra al nebulizador. Existen varios nebulizadores disponibles comercialmente, entre los que se encuentran los neumáticos, ultrasónicos y desolvatantes. Los nebulizadores neumáticos que usan flujo de gas para generar el aerosol son el tipo más común utilizado. Dentro de los nebulizadores neumáticos, existen varios tipos diferentes, los

concéntricos, de flujo cruzado, Babington y ranura en V (una variante del Babington). Cada uno tiene sus propias ventajas y desventajas; la elección de uno u otro depende de la aplicación. Los nebulizadores concéntricos son ideales para muestras de matriz simple (sólidos disueltos totales (TDS) bajos), mientras que los nebulizadores de flujo cruzado y ranura en V son más resistentes y toleran muestras de matriz compleja. A diferencia del diseño neumático, un nebulizador ultrasónico utiliza energía acústica para generar el aerosol. Estos nebulizadores mejoran la sensibilidad analítica en un orden de magnitud en comparación con los nebulizadores neumáticos, sin embargo, son significativamente más caros. Los sistemas nebulizadores desolvatantes utilizan una cámara de pulverización calentada para desolvatar la muestra antes de que llegue al plasma [16].

Después de que la muestra es transformada en un aerosol mediante su paso por el nebulizador, la muestra ingresa a la cámara de nebulización. Esta cámara tiene un diseño simple pero sirve para algunos propósitos importantes: filtra selectivamente las gotas de aerosol más grandes que genera el nebulizador y actúa para suavizar los "pulsos" de nebulización producidos por la bomba peristáltica. Esto es importante porque el plasma es ineficiente para trabajar con gotas grandes (> 10 μm de diámetro).

Generación de iones

Después de que la muestra se nebuliza, el llamado aerosol terciario que sale de la cámara de nebulización (que comprende las gotas más pequeñas del aerosol primario generado por el nebulizador) es barrido en una corriente de gas argón a lo largo del inyector y dentro del plasma. Luego de alcanzar el plasma a alta temperatura, la muestra se desolvata, vaporiza, atomiza e ioniza. La mayoría de los elementos forman iones positivos con carga individual, sin embargo, algunos elementos también pueden formar una pequeña fracción de iones con carga doble. Un ICP típico a presión atmosférica se forma cuando un gas inerte, generalmente Ar, es introducido en una antorcha de cuarzo (que consiste en tres tubos concéntricos de diámetros variables), rodeada de una bobina de inducción conectada a un

generador de radiofrecuencias. Por el canal central de dicha antorcha se arrastra el aerosol de la muestra mediante el gas nebulizador, a un flujo comprendido entre 0,9 y 1,5 L/min. Por el canal más externo de la antorcha circula, a una velocidad de flujo de 15 L/min y de forma tangencial, un corriente de argón que se denomina gas plasma, es decir, aquel que forma el plasma y refrigera la antorcha, ya que previene el contacto físico entre el plasma y las paredes externas de la antorcha, evitando que éstas se fundan. El tubo intermedio de la antorcha lleva un flujo de Ar a 1 L/min, el cual se conoce como gas auxiliar, y es empleado también para refrigerar, ya que mantiene el plasma formado alejado de la boca de la antorcha (Figura 1.13). De esta manera, y por efecto de los flujos de gas plasma y auxiliar, el plasma formado permanece sin contacto con la antorcha y suspendido dentro de la misma.

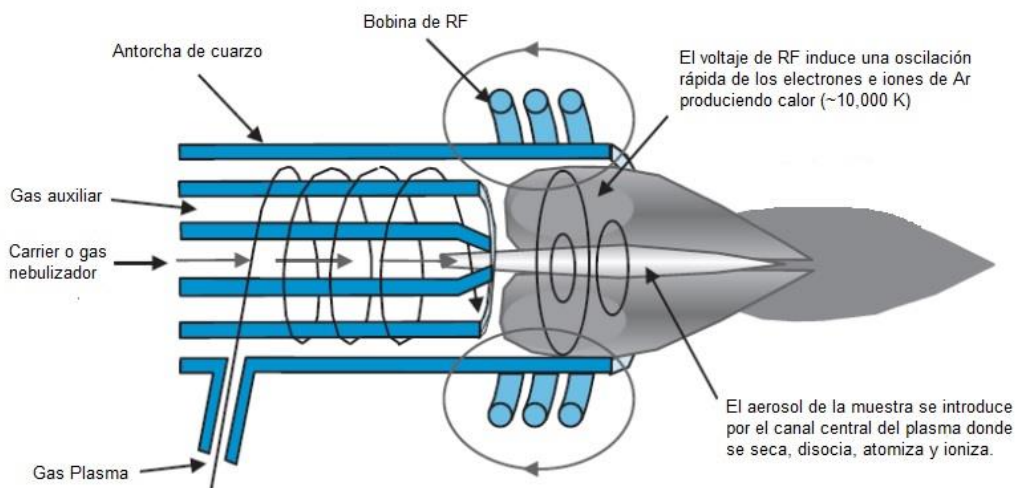


Figura 1.13. Diagrama esquemático de un ICP.

Interfase

En esta etapa, la mezcla de iones y átomos generados luego de la introducción de la muestra dentro del plasma, es impulsada por argón a través de una interface compuesta por dos conos posteriores que permiten la reducción de la presión externa y focalizan los iones en un pequeño volumen para conducirlos hacia el interior del analizador de masas [1].

El primer cono (en contacto con el plasma) se denomina cono de muestreo, y el segundo se llama cono skimmer. Los iones, fotones y átomos o moléculas neutrales se extraen del plasma a la región de interfaz a través de un pequeño orificio en la punta del cono de

muestra (1 mm de diámetro). Se utiliza una bomba mecánica de desbaste o bomba de vacío para mantener una presión de interfaz (entre los conos) de 150–300 Pa [16]. A medida que los iones entran en esta región de interfaz, la reducción dramática de la presión provoca una expansión supersónica de los iones, generando el llamado chorro libre. Posteriormente, se extraen los iones a través de un orificio aún más pequeño en el cono del skimmer (aproximadamente de 0,45 mm de diámetro), y dentro de la cámara de vacío principal que se mantiene bajo alto vacío mediante una bomba turbomolecular. Este cono es la interfase entre la cámara de expansión y la cámara que contiene las lentes. A esta presión, los iones pueden ser guiados efectivamente por superficies cargadas llamadas lentes electrostáticas. Un fluido refrigerante circula continuamente entre una unidad enfriadora y el instrumento (particularmente la bobina de RF y los conos de interfaz) para evitar que estos componentes se sobrecalienten [13,16].

Focalización de iones

El conjunto de lentes electrostáticas ubicadas detrás del cono del skimmer se denomina colectivamente óptica de iones. La función del sistema de óptica iónica es guiar el haz de iones hacia el analizador de masas y evitar que los fotones y otras especies neutras (como los componentes de la matriz no ionizada) lleguen al detector.

Si bien los fotones son la base de la medición en técnicas de emisión atómica como la fotometría de llama y el ICP-AES, en ICP-MS son una fuente de ruido e inestabilidad de la señal si se les permite llegar al detector. La óptica iónica generalmente evita que esto ocurra colocando un "tope de fotones" en la ruta de iones, o guiando el haz de iones fuera del eje.

Como mejoramiento al sistema de lentes iónicas descrito, y en algunas versiones instrumentales, el ICP-MS puede incluir una celda de colisión o reacción llamada "celda universal". Se trata de dispositivos opcionales en los equipos de ICP-MS pero de suma importancia para la eliminación de interferencias poliatómicas en el análisis de matrices complejas. En el caso de la celda de colisión, se hace pasar el haz de iones por un gas de baja

masa atómica (helio) a baja presión. Los iones del analito no se ven afectados por la presencia de este gas, mientras que las moléculas poliatómicas son fragmentadas, convirtiéndose en iones que no interfieren con el analito de interés. En el caso de la celda de reacción, se introduce uno o varios gases altamente reactivos en la celda para favorecer reacciones químicas predecibles. Un cuadrupolo de barrido quita de forma instantánea cualquier nueva interferencia resultante de cualquier reacción secundaria [14].

Analizador de masas

El espectrómetro de masas separa los iones cargados de cada isótopo de acuerdo a su relación masa/carga y, cómo la mayoría de los iones generados en el ICP son monocargados, se indica que el analizador los separa en función de su masa. Se utilizan tres tipos principales de espectrómetros de masas en los equipos comerciales de ICP-MS: cuadrupolo, tiempo de vuelo y sector magnético. Entre los analizadores de masa, el analizador de masas cuadrupolar es más compacto, más barato y más robusto que cualquier otro tipo de espectrómetro de masas, por lo que es el mayormente usado en los equipos de ICP-MS. El analizador de masas cuadrupolar está formado por cuatro barras cilíndricas que actúan como electrodos, por lo general son de molibdeno y están colocadas paralelamente de forma equidistante a un eje central imaginario. Las barras opuestas se conectan eléctricamente, un par está unido al polo positivo de una fuente variable de corriente continua y el otro par se une al terminal negativo. Potenciales de corriente alterna (CA) y corriente continua (CC) de radiofrecuencia se aplican a las barras, creando un campo eléctrico variable en el tiempo en el centro a través del cual pasan los iones. Para un ion con una relación m/z particular, solo combinaciones específicas de potenciales de CA y CC dan como resultado una trayectoria de vuelo de iones estable a través del cuadrupolo. Los iones con trayectorias inestables chocan con las barras que tienen carga opuesta, neutralizándose, por lo que no se transmiten a través del cuadrupolo y no alcanzan el detector. Estos voltajes se pueden aumentar muy rápidamente, lo que permite escanear todo el rango de masas en cuestión de milisegundos. La alta velocidad de exploración de los instrumentos

modernos permite la adquisición de datos para múltiples elementos prácticamente de forma simultánea [14,16].

El período de tiempo que el instrumento pasa adquiriendo datos a un valor m/z particular se denomina tiempo de permanencia (*dwell time* en inglés). Los tiempos de permanencia más largos permiten mediciones más precisas (y sensibles) en el detector, promediando la señal durante un período de tiempo más largo. Otros parámetros del cuadrupolo como el modo de adquisición de datos (escaneo continuo o salto de pico) y el número de escaneos o réplicas también pueden definirse mediante el analista [16].

Detector

El detector más comúnmente utilizado para ICP-MS es un multiplicador de electrones (EM). Los iones de analito cargados positivamente golpean el primer dinodo del detector que se mantiene a un alto voltaje negativo. El impacto del ion en el detector provoca la emisión de varios electrones desde la superficie, que, a su vez, golpean el próximo dinodo liberando más electrones. Este proceso (llamado emisión secundaria) continúa, generando una cascada de amplificación que culmina en una señal lo suficientemente grande como para medirse de manera confiable como un "recuento" de iones [16].

1.3.3. Referencias

- [1] M. Aceto, The Use of ICP-MS in Food Traceability, in: *Adv. Food Traceability Tech. Technol.*, Elsevier, 2016: pp. 137–164. doi:10.1016/B978-0-08-100310-7.00008-9.
- [2] G.P. Danezis, A.S. Tsagkaris, F. Camin, V. Brusica, C.A. Georgiou, Food authentication: Techniques, trends & emerging approaches, *TrAC - Trends Anal. Chem.* 85 (2016) 123–132. doi:10.1016/j.trac.2016.02.026.
- [3] A. Gonzalez, S. Armenta, M. de la Guardia, Trace-element composition and stable-isotope ratio for discrimination of foods with Protected Designation of Origin, *TrAC Trends Anal. Chem.* 28 (2009) 1295–1311. doi:10.1016/j.trac.2009.08.001.
- [4] A. González, M. de la Guardia, Mineral profile, in: *Compr. Anal. Chem.*, 2013: pp. 51–

76. doi:10.1016/B978-0-444-59562-1.00003-7.
- [5] S. Esslinger, J. Riedl, C. Fauhl-Hassek, Potential and limitations of non-targeted fingerprinting for authentication of food in official control, *Food Res. Int.* 60 (2014) 189–204. doi:10.1016/j.foodres.2013.10.015.
- [6] S. Medina, J.A. Pereira, P. Silva, R. Perestrelo, J.S. Câmara, Food fingerprints – A valuable tool to monitor food authenticity and safety, *Food Chem.* 278 (2019) 144–162. doi:10.1016/j.foodchem.2018.11.046.
- [7] S. Armenta, M. de la Guardia, Analytical Approaches for the Evaluation of Food Protected Designation of Origin, in: *Adv. Food Traceability Tech. Technol.*, Elsevier, 2016: pp. 275–301. doi:10.1016/B978-0-08-100310-7.00015-6.
- [8] D. Cozzolino, S. Heather, Analytical and chemometric-based methods to monitor and evaluate wine protected designation, 1st ed., Copyright © 2013 Elsevier B.V. All rights reserved., 2013. doi:10.1016/B978-0-444-59562-1.00015-3.
- [9] J. Saurina, Characterization of wines using compositional profiles and chemometrics, *TrAC - Trends Anal. Chem.* 29 (2010) 234–245. doi:10.1016/j.trac.2009.11.008.
- [10] P. Pohl, What do metals tell us about wine?, *TrAC - Trends Anal. Chem.* 26 (2007) 941–949. doi:10.1016/j.trac.2007.07.005.
- [11] B. Tariba, Metals in Wine—Impact on Wine Quality and Health Outcomes, *Biol. Trace Elem. Res.* 144 (2011) 143–156. doi:10.1007/s12011-011-9052-7.
- [12] D.M.A.M. Luykx, S.M. van Ruth, An overview of analytical methods for determining the geographical origin of food products, *Food Chem.* 107 (2008) 897–911. doi:10.1016/j.foodchem.2007.09.038.
- [13] D.A. Skoog, J.F. Holler, S.R. Crouch, *Principles of Instrumental Analysis*, sixth edit, Thomson Brooks/Cole, Belmont, CA., 2007.
- [14] Perkin Elmer Instruments, *ICP-Mass Spectrometry. The 30-Minute Guide to ICP-MS*, (2004) 8.

- [15] D. Beauchemin, Inductively Coupled Plasma Mass Spectrometry Methods, in: *Encycl. Spectrosc. Spectrom.*, 3rd ed., Elsevier, 2017: pp. 236–245. doi:10.1016/B978-0-12-409547-2.11222-3.
- [16] S.C. Wilschefski, M.R. Baxter, Inductively Coupled Plasma Mass Spectrometry: Introduction to Analytical Aspects, *Clin. Biochem. Rev.* 40 (2019) 115–133. doi:10.33176/AACB-19-00024.

1.4. ANÁLISIS QUIMIOMÉTRICO PARA LA DESIGNACIÓN DE ORIGEN

1.4.1. Quimiometría e introducción al análisis multivariado

La quimiometría se puede definir como la disciplina dentro de la química que, basándose en la aplicación de herramientas matemáticas y estadísticas, desarrolla métodos para obtener información relevante de los datos químicos [1,2]. Estas técnicas se consideran herramientas útiles para la evaluación de la autenticidad y la trazabilidad, así como para la detección de cualquier adulteración de alimentos, entre otras aplicaciones importantes.

La quimiometría nace como disciplina científica en la década de los años 70 del siglo pasado, con el desarrollo de la instrumentación y el gran auge de la informática. La palabra “quimiometría” data su origen en el año 1972 introducida por Svante Wold y Bruce R. Kowalski, quienes fundaron la “International Chemometrics Society” en 1974 y con ello condujeron a la primera descripción formal de esta disciplina [3].

Los métodos quimiométricos se aplican sobre grandes cantidades de datos de características multivariantes, es decir que una caracterización completa de las muestras implica la determinación de muchas características analíticas que, por lo tanto, requieren el despliegue de muchas técnicas analíticas diferentes. Tales características deben considerarse juntas, porque sus interrelaciones tienen un papel crucial en la caracterización, para proporcionar lo que se llama una huella dactilar analítica o fingerprint para la autenticidad de los productos. Por esta razón, generalmente se requiere un enfoque de reconocimiento de patrones multivariados para obtener modelos exitosos para verificar las afirmaciones de PDO [4]. La aplicación del análisis de datos multivariados es una herramienta poderosa para analizar y estructurar conjuntos de datos que se han obtenido, y para hacer modelos matemáticos empíricos que son, por ejemplo, capaces de predecir los valores de propiedades importantes.

La quimiometría entonces engloba una gran cantidad de datos obtenidos por el análisis de múltiples variables sobre una gran variedad de muestras, atendiendo a esto es que aparece en la literatura el concepto de minería de datos o data mining y el de aprendizaje automático o

machine learning. Aunque ambos términos se usan indistintamente para referirse a las técnicas de procesamiento de datos, existen diferencias entre ellas. La minería de datos se puede utilizar con fines descriptivos (es decir, mostrar similitudes entre los elementos de un conjunto de datos) o con fines predictivos (es decir, predecir características específicas de nuevos datos basados en modelos que se han construido y validado previamente). Se basa en la recopilación, almacenamiento y tratamiento de una gran cantidad de datos para tomar las mejores decisiones sobre un problema en particular. Las herramientas complejas de tratamiento de datos se utilizan para detectar e identificar patrones ocultos, asociaciones y estructuras que son propias del conjunto de datos grandes sin procesar, o para seleccionar y filtrar información útil de grandes bases de datos. Por otro lado, el concepto de aprendizaje automático también se conoce como las técnicas involucradas en el manejo de vastos datos de la manera más inteligente (mediante el desarrollo de algoritmos) para obtener información procesable. En estas técnicas, esperamos que los algoritmos aprendan por ellos sin ser programados explícitamente. En consecuencia, la minería de datos se refiere al área en general y el aprendizaje automático se refiere exclusivamente a los algoritmos utilizados y está vinculado al reconocimiento de patrones [5].

En los últimos años, la minería de datos se ha utilizado con mayor frecuencia en el área de análisis de alimentos, dejando los conceptos de técnicas o métodos de reconocimiento de patrones y herramientas quimiométricas para referirse a los algoritmos utilizados para procesar los datos. Tanto la minería de datos como la quimiometría representan conceptos muy similares. De hecho, la única diferencia es que la quimiometría se ha utilizado en referencia a la aplicación de técnicas de aprendizaje automático para obtener información del sistema material a partir de datos de naturaleza principalmente química o físico-química, mientras que la minería de datos se usa ampliamente en muchas otras áreas tales como seguridad, reconocimiento facial, marketing personalizado, diagnóstico médico, navegación aérea, etc. [6].

En esta sección del capítulo se explicarán las etapas del análisis quimiométrico que incluyen los pasos relacionados con la transformación y adaptación de los datos para su posterior análisis, el análisis exploratorio de los datos y finalmente el análisis clasificatorio.

1.4.1.1. Datos multivariados

Como se dijo anteriormente, la quimiometría es una disciplina puramente multivariada. Los métodos y modelos del análisis multivariado (AM) estudian, analizan, representan e interpretan los datos que resultan de observar más de una variable estadística sobre una muestra de individuos. El objetivo del análisis de datos multivariados es determinar todas las variaciones en el estudio de una matriz de datos. La información estadística en AM es de carácter multidimensional, por lo tanto la geometría, el cálculo matricial y las distribuciones multivariantes juegan un papel fundamental [7].

La información multivariante se dispone generalmente en una matriz de datos, de dimensión $n \times p$, siendo n el número de casos u objetos (muestras) y p el número de variables registradas sobre cada caso. Llamaremos X , a la matriz de datos de dimensión $n \times p$ donde cada fila es un caso u observación multivariada. Una observación multivariada es la colección de mediciones sobre p variables diferentes tomadas sobre el mismo ítem o unidad objeto de estudio (Fig. 1.14) [8].

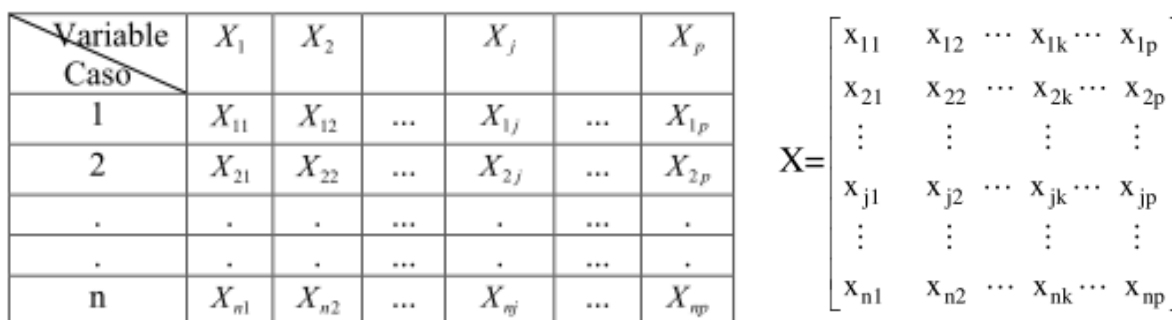


Figura 1.14. Organización de los datos multivariados.

Cada observación multivariada es representada por un vector p-dimensional de variables aleatorias y puede ser conceptualizada como un punto (o vector desde el origen) en

el espacio p-dimensional con coordenadas igual al valor de cada una de las variables para esa observación. El vector de variables para una observación puede considerarse como una variable vectorial o multivariante p-variada [8]. Casi todas las técnicas del análisis multivariado están basadas en el concepto de distancia. La distancia entre un par de puntos que representan casos en el espacio que tiene como ejes a las variables indica la disimilitud entre los casos. A mayor distancia, menor parecido o similitud multivariada entre los casos. Bajo el concepto de distancia, los agrupamientos de objetos y datos extremos presentes pueden ser detectados.

1.4.1.2. Pre-procesamiento de datos

Uno de los aspectos importantes a la hora de aplicar las herramientas estadísticas de análisis multivariado es el procesamiento previo de los datos. Usualmente los datos se presentan en formatos no óptimos o incluso inadecuados para ser procesados por los modelos. En estos casos el pre-procesamiento de los datos es un paso obligatorio para la calidad de los resultados. Muchos algoritmos requieren que las características estén en la misma escala (por ejemplo en el rango [0,1]) para optimizar su rendimiento, lo que se realiza aplicando técnicas de normalización o estandarización de los datos. En otros casos, podemos también encontrar que las características o variables seleccionadas están correlacionadas, y por lo tanto son redundantes para extraer información. Entonces, se usan técnicas de reducción dimensional para comprimir las características en subespacios con menores dimensiones. Los pasos involucrados en el pre-procesamiento de datos implican, limpieza, integración, reducción y transformación de los datos [9].

Las rutinas de limpieza de datos trabajan para "limpiar" los datos rellenando los valores faltantes, suavizando los datos ruidosos, identificando o eliminando valores atípicos y resolviendo inconsistencias. En el caso que se incluyan datos de múltiples fuentes en el análisis, es necesario entonces la integración de múltiples bases de datos, o archivos, es decir, la integración de datos. Con la reducción de datos se obtiene una representación reducida del

conjunto de datos que es mucho menor en volumen, pero produce los mismos (o casi los mismos) resultados analíticos. Las estrategias de reducción de datos incluyen reducción de dimensionalidad y reducción de numerosidad. En la reducción de dimensionalidad, se aplican esquemas de codificación de datos para obtener una representación reducida o "comprimida" de los datos originales. Los ejemplos incluyen técnicas de compresión de datos (por ejemplo, análisis de componentes principales), selección de un subconjunto de atributos (por ejemplo, eliminación de atributos irrelevantes) y construcción de atributos (por ejemplo, donde se deriva un pequeño conjunto de atributos más útiles del conjunto original). En la reducción de numerosidad, los datos se reemplazan por representaciones alternativas más pequeñas utilizando modelos paramétricos (por ejemplo, modelos de regresión o log-lineales) o modelos no paramétricos (por ejemplo, histogramas, agrupaciones, muestreo o agregación de datos). Finalmente la transformación de los datos incluye principalmente estrategias matemáticas que convierten los datos en las formas apropiadas para la posterior aplicación de los métodos. La unidad de medida utilizada puede afectar el análisis de los datos, por ejemplo, cambiar las unidades de medida de metros a pulgadas para la altura, o de kilogramos a libras para el peso, puede conducir a resultados muy diferentes. En general, la expresión de un atributo en unidades más pequeñas dará lugar a un rango mayor para ese atributo y, por lo tanto, tenderá a darle a dicho atributo un mayor efecto o "peso". Para ayudar a evitar la dependencia de la elección de las unidades de medida, los datos deben normalizarse o estandarizarse. Esto implica transformar los datos para que caigan dentro de un rango más pequeño o común, como $[-1; 1]$ o $[0,0; 1,0]$. La normalización de los datos intenta dar a todos los atributos un mismo peso. La normalización es particularmente útil para algoritmos de clasificación que involucran redes neuronales o mediciones de distancia, tales como la clasificación del k-vecino más cercano y clustering o análisis de conglomerados [9].

Existen diversos métodos para la normalización de datos, entre ellos, la normalización min-max, la normalización de transformación z y la normalización por escala decimal.

La *normalización mínima-máxima* realiza una transformación lineal en los datos originales. Si min_A y max_A son los valores mínimos y máximos de un atributo, A. La normalización min-max asigna un valor, v_i de A a v'_i en el rango $[nueva_min_A, nueva_max_A]$ calculando:

$$v'_i = \frac{v_i - min_A}{max_A - min_A} (nueva_max_A - nueva_min_A) + nueva_min_A$$

En la *normalización de la transformación z* (normalización de media cero o autoescalado), los valores para un atributo A, se normalizan en función de la media (es decir, el promedio) y la desviación estándar de A. Un valor, v_i , de A se normaliza a v'_i por calculando:

$$v'_i = \frac{v_i - \bar{A}}{\sigma_A},$$

donde \bar{A} y σ_A son la media y la desviación estándar, respectivamente, del atributo A.

En la *normalización por escala decimal* se normaliza moviendo el punto decimal de los valores del atributo A. El número de puntos decimales movidos depende del valor absoluto máximo de A. Un valor, v_i , de A se normaliza a v'_i al calcular:

$$v'_i = \frac{v_i}{10^j},$$

donde j es el entero más pequeño tal que $\max(|v'_i|) < 1$.

Las técnicas de pre-procesamiento de datos pueden mejorar la calidad de los datos, lo que ayuda a mejorar la precisión y la eficiencia del proceso posterior aplicado. El pre-procesamiento de datos es un paso importante en el proceso de descubrimiento de conocimiento, porque las decisiones de calidad deben basarse en datos de calidad [9]. A modo de corolario, la siguiente figura esquematiza las formas de pre-procesamiento de datos (Fig. 1.15).

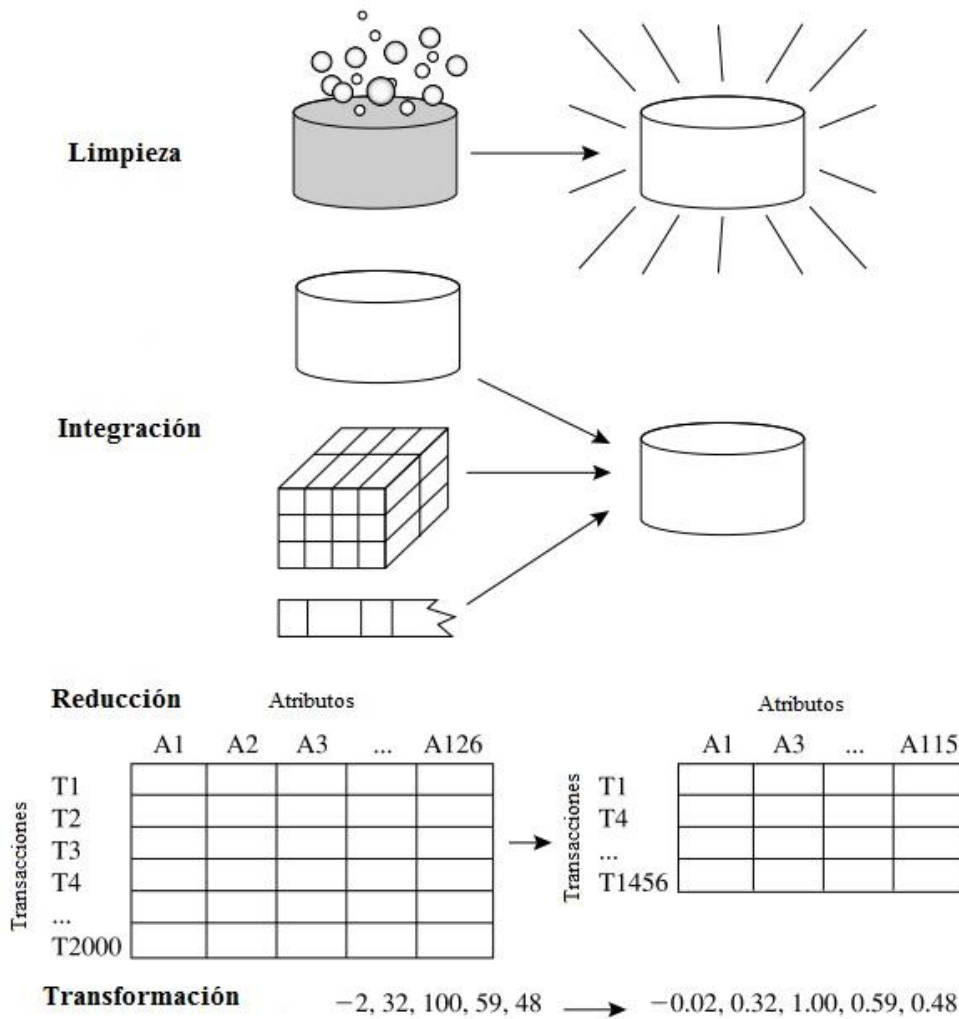


Figura 1.15. Técnicas de pre-procesamiento de los datos [9].

1.4.2. Análisis exploratorio de datos

El análisis exploratorio de datos (EDA) es conocido como el enfoque operativo para el análisis de datos destinado a mejorar la comprensión y la accesibilidad de los resultados. EDA se relaciona con el proceso de revelar información oculta y desconocida de los datos de tal manera que el analista obtiene una representación inmediata, directa y fácil de entender. Los gráficos visuales son un elemento de este enfoque, debido a la capacidad intrínseca del cerebro humano para obtener una interpretación más directa y confiable de similitudes, diferencias, tendencias, agrupaciones y correlaciones a través de una imagen, en lugar de una serie de números [10]. Por lo tanto, las herramientas y técnicas de EDA están fuertemente

basadas en el enfoque gráfico. La visualización de datos se realiza mediante diagramas de caja, histogramas y diagramas de dispersión, todos los instrumentos libres de distribución que pueden ser extremadamente útiles para investigar si los datos siguen una distribución particular.

A la hora de la existencia de datos multivariados, se persiguen los mismos objetivos, a saber, dar una representación gráfica de datos multivariados destacando los patrones de datos y las relaciones entre objetos y variables sin una formulación de hipótesis a priori. Se necesitan herramientas de detección multivariadas para modelar los factores funcionales latentes subyacentes que determinan lo que sucede en los sistemas examinados y son la base de una estrategia exploratoria de datos inductivos [10].

Estas herramientas tienen que cumplir dos tareas:

- compresión de toda la información a un pequeño conjunto de parámetros sin introducir distorsión de la estructura de datos (o al menos mantenerla al mínimo y mantener el control de la perturbación que se ha introducido);
- representación gráfica eficiente de los datos.

Las técnicas más efectivas para lograr estos objetivos se basan en técnicas de proyección, es decir, métodos para proyectar los datos desde su espacio de J -variables/condiciones a una dimensionalidad más baja, es decir, espacio de A -factores latentes/componentes. Es entonces que el EDA consiste principalmente en dos técnicas: Análisis de Componentes Principales (PCA) y Análisis Factorial (FA) [11].

El método más común para el análisis exploratorio de datos es el PCA. Sin embargo, si el número de variables es pequeño para el caso de estudio considerado, también las estadísticas univariadas simples pueden ser informativas [12].

1.4.2.1. Análisis de Componentes Principales (PCA)

PCA es una técnica multivariada bien conocida para el análisis exploratorio de datos, que proyecta los datos en un hiperespacio reducido, definido por los componentes principales.

Estas son combinaciones lineales de las variables originales, con el primer componente principal con la mayor varianza explicada, el segundo componente principal con la segunda mayor varianza, y así sucesivamente [13].

El análisis de componentes principales puede ser considerado como “la madre de todos los métodos en el análisis de datos multivariados” [14]. El objetivo de PCA es la reducción de dimensiones y es el método más frecuentemente aplicado para el cálculo de variables latentes lineales (componentes). PCA puede ser visto como un método para calcular un nuevo sistema de coordenadas formado por variables latentes, que son ortogonales. Estas primeras variables son las que más información aportan de todo el set de variables. Las variables latentes del PCA representan de manera óptima las distancias entre los objetos en el espacio de altas dimensiones de las variables. PCA considera todas las variables y acomoda la estructura total de los datos; se trata de un método exploratorio de análisis de datos (aprendizaje no supervisado) y se puede aplicar a cualquier matriz X [14].

La reducción de dimensiones por PCA es empleada principalmente para:

- Visualización de los datos multivariados por gráficos de dispersión.
- Transformación de variables altamente correlacionadas en variables no correlacionadas que pueden ser usadas por otros métodos.
- Separación de información relevante (descrito por unas pocas variables latentes) del ruido.
- Combinación de varias variables que caracterizan un proceso químico-tecnológico en una o pocas variables.

El PCA es un método exitoso para un conjunto de datos con variables correlacionadas como es usual en datos químicos. Las variables constantes o variables altamente correlacionadas no causan problemas para el PCA; sin embargo, los outliers o datos extremos influyen en los resultados, entonces el escalado se vuelve importante.

En el PCA clásico, la matriz de datos original se modifica por un pretratamiento basado en el centrado de los resultados experimentales y, a veces, en su escalado, para garantizar que el primer componente principal describa la dirección correcta de la varianza máxima. Después del tratamiento previo de los datos, la matriz de datos X de dimensiones " $m \times n$ ", siendo m el número de variables medidas y n el número de objetos, se descompone en un producto de matriz de scores (valores proyectados en las variables latente), con " $m \times a$ " (siendo a el número de PCA) y la matriz de loadings que tiene dimensiones " $a \times n$ ". Se presenta la siguiente ecuación:

$$X=TP^T$$

donde T es la matriz de scores y P^T la matriz transpuesta de loadings.

Sin embargo, debe tenerse en cuenta que un conjunto reducido de PCA no puede explicar completamente la varianza de los datos experimentales y que existe una relación entre las matrices antes mencionadas, que incluye una matriz de residuos que debe minimizarse incorporando PC adicionales.

$$X=TP^T+E$$

Entonces, como se observa en la Fig. 1.16, la matriz de variables medidas originalmente, obtenida para las n muestras consideradas (objetos), proporciona un espacio de dimensión m , que define la posición de cada muestra. La fuerte reducción de este espacio de m a un conjunto de componentes principales, a , crea un nuevo posicionamiento de los objetos en el nuevo espacio ortogonal representado por la matriz de scores y, por lo tanto, la orientación de los nuevos ejes con respecto a las coordenadas originales, define la matriz de loadings, que es una matriz " $a \times m$ ", explicando la varianza total de los datos al aumentar el número de PC, siendo la matriz de residuos, más pequeña [15].

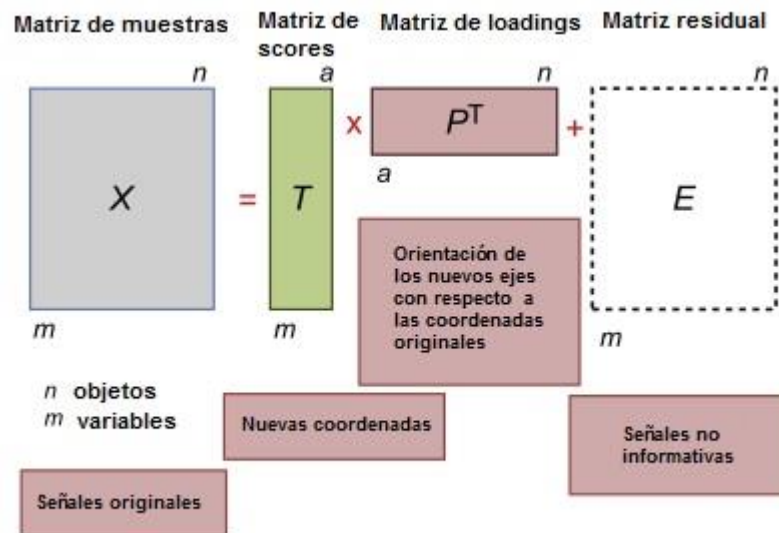


Figura 1.16. Esquema del proceso de reducción de variables en PCA [15].

Los vectores de scores, proporciona las coordenadas de las muestras en el espacio de la PC, por lo tanto, los gráficos de dispersión de los scores permiten la inspección de la similitud/disimilitud de la muestra y los vectores de loadings, representan el peso con el que cada variable original contribuye a las PC, de modo que la estructura de correlación entre las variables se puede inspeccionar mediante de gráficos de dispersión de loadings. E es la matriz residual, ruido o error, la parte de los datos que no fue explicada por el modelo; tiene las mismas dimensiones que X y a menudo se usa como herramienta de diagnóstico para la identificación de muestras y/o variables outliers [10].

La reducción antes mencionada en las coordenadas ha demostrado ser de gran valor como pretratamiento para cálculos complejos, y la reducción de hiperespacios a planos y volúmenes permite una interpretación de las relaciones entre muestras y la interpretación del significado de sus variables retenidas. De hecho, el espacio reducido de PCA proporciona nuevas relaciones entre los objetos considerados, lo que permite identificar la existencia de diferentes grupos.

Entonces la ventaja de este método radica en que muestra rápida y fácilmente similitudes y diferencias entre las muestras y las relaciones de las variables, sin embargo no permite clasificar y asignar una clase a cada muestra.

1.4.2.2. Análisis Factorial (FA)

A diferencia del PCA, que puede considerarse como un método para la rotación de bases, el análisis factorial se basa en un modelo estadístico con ciertos supuestos del modelo. Al igual que la PCA, el análisis factorial también da como resultado una reducción de la dimensión, pero si bien la PCA solo se deriva de la optimización de un criterio estadístico (dispersión, varianza), los factores están destinados a tener un significado real e interpretaciones [14].

FA explica la variabilidad entre las variables aleatorias observadas en términos de algunas variables aleatorias no observadas y un método de detección de estructura que revela el patrón de relaciones variables. Las variables independientes latentes inferidas se denominan factores. Las variables observadas se modelan como combinaciones lineales de los factores y términos de error. Los factores extraídos consecutivos explican cada vez menos variabilidad a medida que avanza el proceso; el criterio de detención se basa en el nivel de variabilidad aleatoria (es decir, error) dejado de lado que puede ser tolerado. La FA se aplica típicamente a una matriz de correlación [16].

1.4.3. Análisis clasificadorio o de reconocimiento de patrones

La clasificación es la operación de separar varias entidades en varias clases. Los métodos de clasificación son técnicas quimiométricas fundamentales diseñadas para encontrar modelos matemáticos capaces de reconocer la pertenencia de cada objeto a su clase apropiada sobre la base de un conjunto de mediciones. Una vez que se ha obtenido un modelo de clasificación, se puede predecir la pertenencia de objetos desconocidos a una de las clases definidas. Las técnicas de clasificación (clasificadores) son métodos cuantitativos para modelar respuestas cualitativas. En otras palabras, los métodos de clasificación encuentran relaciones matemáticas entre un conjunto de variables descriptivas (por ejemplo, mediciones químicas) y un método cualitativo. [13,17].

Cuando se habla de los métodos de clasificación aparece un término muy empleado, mayormente en el campo de la computación, son las técnicas de reconocimiento de patrones. El reconocimiento de patrones tiene como objetivo clasificar los datos basados en el conocimiento a priori o en la información estadística extraída de los patrones. Es una disciplina científica cuyo objetivo es no sólo la clasificación de objetos en varias categorías o clases, sino también la posibilidad de determinar qué tan bien se puede hacer esta clasificación y así validar el modelo. Además, puede determinar cuántas clases y también si hay valores atípicos y qué características son mejores para distinguir los grupos. Entonces, aunque los algoritmos de clasificación están en el centro de la mayoría de los métodos de reconocimiento de patrones, no son exclusivos [18].

Como se observa en la Fig. 1.17, los métodos de reconocimiento de patrones se pueden dividir en dos tipos. Los enfoques supervisados intentan dividir los objetos en grupos según sus características utilizando un conjunto de entrenamiento, es decir, objetos que están etiquetados en categorías predefinidas. La clasificación predice una categoría, mientras que la regresión predice un número. Los enfoques no supervisados por su parte, intentan dividir el espacio de datos en grupos sin ningún conjunto de entrenamiento predefinido, por lo que el enfoque de clasificación no supervisado más común es la agrupación [17,18].



Figura 1.17. Enfoques de los métodos de reconocimiento de patrones.

El desarrollo de los métodos quimiométricos de clasificación (Fig. 1.18) implica diferentes pasos, el primero de ellos consiste en configurar los resultados experimentales recopilados en un formato de matriz, X. Posteriormente, puede ser necesario hacer un pre-procesamiento de datos, que dependerá de la naturaleza de ellos. Una vez obtenida la data transformada con los métodos exploratorios multivariados se obtiene una primera visualización de la estructura de los datos.

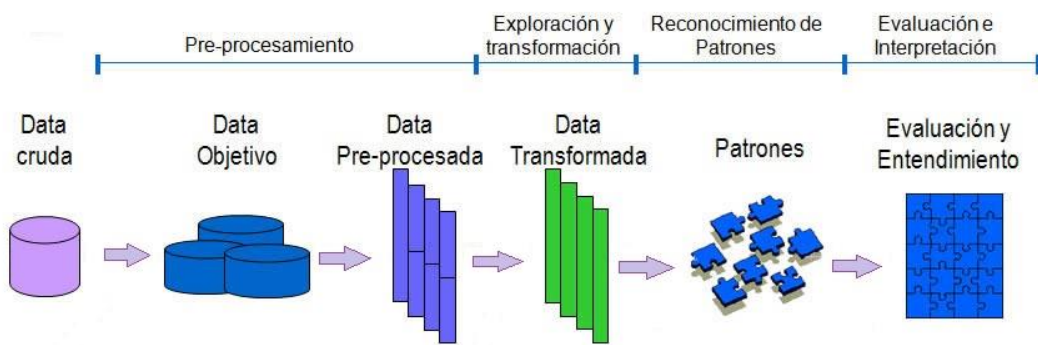


Figura 1.18. Proceso de clasificación [5].

El desarrollo de un modelo supervisado de reconocimiento de patrones implica dos etapas. La primera etapa es construir el modelo estimando los parámetros óptimos, utilizando un conjunto de objetos o muestras cuya clase o características particulares son conocidas (es decir, conjunto de entrenamiento o conjunto de calibración). En esta etapa, se podría aplicar una validación interna o una validación cruzada para evaluar la bondad del ajuste del modelo a partir de las muestras/objetos del conjunto de entrenamiento. La segunda etapa es evaluar y validar externamente el rendimiento del modelo construido en la etapa anterior; esto se realiza utilizando objetos o muestras adicionales (es decir, un conjunto de prueba o un conjunto de validación) que cumplen los mismos requisitos pero que no formaban parte del conjunto de entrenamiento original. En estos métodos, se supone que hay suficientes objetos/muestras de referencia que actúan como estándares analíticos porque los resultados

de interés (es decir, la clase cualitativa o el valor de una o más características cuantitativas) se conocen anteriormente o se han medido con precisión [6,19,20].

1.4.3.1. Métodos no supervisados

Como se expuso anteriormente, el aprendizaje no supervisado tiene lugar cuando no se dispone de datos “etiquetados” para el entrenamiento. Por tanto, sólo se puede describir la estructura de los datos, para intentar encontrar algún tipo de organización que simplifique el análisis. Por ello, tienen un carácter exploratorio.

Los tipos de algoritmos más habituales en aprendizaje no supervisado son:

- Clustering o análisis de conglomerados (CA)
- Análisis de componentes principales (PCA)

El reconocimiento de patrones no supervisados incluye algoritmos que pueden emplearse para realizar un análisis exploratorio como es el caso del PCA, sin embargo difiere del análisis exploratorio de datos en que el objetivo de los métodos es detectar similitudes, mientras que el uso de EDA no tiene ningún prejuicio particular sobre si se encontrarán o no grupos y cuántos [11].

A continuación sólo se detallará sobre el análisis de conglomerados o clúster ya que el de componentes principales se expuso anteriormente (punto 1.4.2.1).

Análisis de conglomerados o clustering (CA)

La agrupación en clúster o conglomerados es uno de los métodos importantes para descubrir el conocimiento en datos multidimensionales. El objetivo de la agrupación es identificar patrones, grupos o conglomerados de objetos similares dentro de un conjunto de datos de interés basado en el principio por el cual objetos similares están representados por puntos cercanos en el espacio de las variables que los describen [21].

La tarea de encontrar grupos similares de objetos presume que esa estructura de grupos está inherente en el conjunto de datos. En general, no se asume que un objeto pertenece a un único grupo, pero puede ser parte de dos o más grupos. Entonces, los métodos

de agrupamiento que realizan una separación de objetos en grupos separados no siempre dan la solución deseada. Por esta razón, muchos algoritmos de conglomerados que se proponen en la literatura no sólo operan de manera distinta, sino que incluso actúan sobre principios diferentes [14].

Los métodos posibles difieren en cómo se definen los grupos o en el algoritmo utilizado para crearlos. En términos generales, la definición de grupo se basa en medidas dentro del grupo (por ejemplo, alta similitud entre observaciones) o, alternativamente, en medidas entre grupos (por ejemplo, la distancia máxima entre objetos), mientras que los algoritmos de agrupamiento se basan en diferentes formas de definir la proximidad, ya sea similitudes o diferencias. La forma más intuitiva de definir el nivel de similitud de las muestras (el concepto de disimilitud es complementario, es decir, su valor aumenta cuanto más diferentes son los objetos, mientras que la similitud aumenta cuanto más similares son los objetos) se basa en la conversión de una matriz $N \times M$ de datos en una matriz $N \times N$ de distancias D obtenidas definiendo una métrica, como la distancia euclidiana:

$$d_{ij} = \sqrt{\sum_{m=1}^M (x_{im} - x_{jm})^2}$$

donde la suma se extiende sobre las M variables que caracterizan cada par de objetos i y j .

Los algoritmos de clustering se pueden dividir en dos familias principales: partición y métodos jerárquicos. El particionamiento tiene como objetivo segmentar un gran conjunto de datos de objetos heterogéneos en k grupos, donde k es conocido a priori o hipotetizado de forma exploratoria o 'descubierto' por el algoritmo de forma iterativa. Por el contrario, la agrupación jerárquica opera en una descomposición anidada en varios niveles de similitud/disimilitud, trabajando de acuerdo con un enfoque ascendente o descendente [10].

La agrupación en particiones clasifica las observaciones, dentro de un conjunto de datos, en múltiples grupos según su similitud. Los algoritmos requieren que el analista especifique el número de clústeres que se generarán. Un método representativo de esta

familia es la agrupación de K -medias, en la que cada clúster está representado por el centro o la media de los puntos de datos que pertenecen a la agrupación. El método K -means es sensible a puntos de datos anómalos y valores atípicos.

La agrupación K -means es el algoritmo de aprendizaje automático no supervisado más utilizado para dividir un conjunto de datos dado en un conjunto de k grupos, donde k representa el número de grupos previamente especificados por el analista. El algoritmo clasifica objetos en múltiples grupos de modo que los objetos dentro del mismo grupo son lo más similares posible (es decir, tiene alta similitud intraclass), mientras que los objetos de diferentes grupos son lo más diferentes posible (es decir, baja similitud entre clases). En la agrupación de k -medias, cada grupo está representado por su centro (centroide) que corresponde a la media de los puntos asignados al grupo. La idea básica detrás de la agrupación de k -means consiste en definir grupos para que la variación total dentro del grupo (conocida como variación total dentro del grupo) se minimice. Los pasos de asignación de clúster y actualización del centroide se repiten iterativamente hasta que las asignaciones de clúster dejan de cambiar (es decir, hasta que se logra la convergencia), es decir, los grupos formados en la iteración actual son los mismos que los obtenidos en la iteración anterior [10,21].

Las técnicas de agrupamiento jerárquico son una segunda categoría importante de los métodos de agrupamiento. Existen dos enfoques básicos para generar una agrupación jerárquica [14]:

- Métodos aglomerativos: en el primer nivel de la jerarquía, cada uno de los n objetos forma un clúster separado, lo que da como resultado n grupos. En el siguiente nivel, los dos grupos más cercanos se fusionan, y así sucesivamente, hasta que finalmente todos los objetos estén en un solo grupo.
- Métodos divisivos: en el primer nivel de la jerarquía, todos los n objetos están en un solo grupo. En el siguiente nivel, este grupo se divide en dos grupos más pequeños. El siguiente

nivel nuevamente divide uno de los grupos en dos grupos más pequeños, y así sucesivamente, hasta que finalmente cada objeto forme un grupo separado.

Dividir un clúster es más exigente desde el punto de vista computacional que fusionar dos clústeres, porque no sólo se debe encontrar el clúster que se va a dividir, sino que también se deben identificar los objetos que formarán dos nuevos clústeres. Por lo tanto, los métodos divisivos no se usan con mucha frecuencia.

La información básica para dividir o fusionar grupos es la similitud o la distancia entre los grupos. Se define $x_i^{(j)}$ y $x_j^{(l)}$ para $i = 1, \dots, n$ como todas las observaciones que se han asignado al grupo j y l , respectivamente, con los tamaños de grupo n_j y n_l . Entonces, la distancia entre los dos grupos con índice j y l se puede determinar por varios métodos: encadenamiento simple, encadenamiento completo, encadenamiento medio, método de Ward y método del centroide [14].

Los clústeres jerárquicos a menudo se muestra gráficamente utilizando un diagrama en forma de árbol llamado *dendograma* (Fig. 1.19), que muestra tanto las relaciones clúster-subclúster como el orden en el que los clústeres se fusionaron (clúster aglomerativo) o se dividieron (clúster divisivo) [22]. Este diagrama en forma de árbol muestra los objetos como "hojas" en la parte inferior, y las ramas se fusionan de acuerdo con el orden dado por el algoritmo.

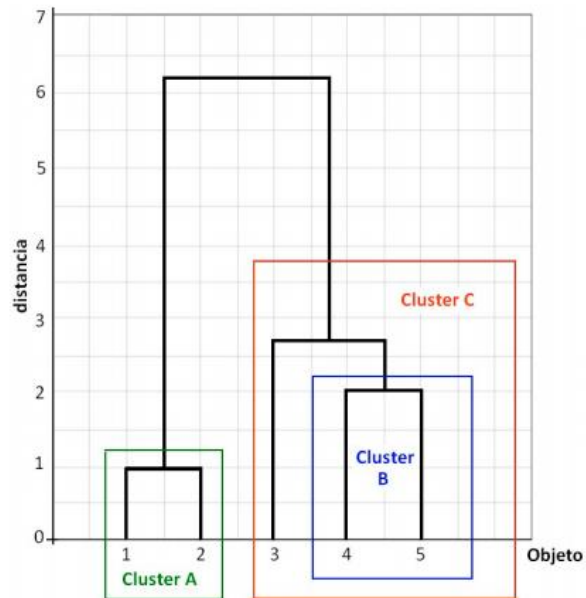


Figura 1.19. Representación ejemplo de un dendrograma.

1.4.3.2. Métodos supervisados

Los métodos supervisados abarcan técnicas que se utilizan para organizar los objetos en diferentes categorías con la ayuda del supervisor, es decir, los algoritmos trabajan con datos “etiquetados”, intentando encontrar una función que, dadas las variables de entrada, les asigne la etiqueta de salida adecuada. El algoritmo se entrena con un “histórico” de datos y así “aprende” a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, predice el valor de salida. El aprendizaje supervisado se llama así porque el desarrollador actúa como una guía para enseñar al algoritmo las conclusiones a las que debe llegar, es decir la salida del algoritmo ya es conocida.

Los métodos de clasificación supervisados se distinguen en la forma matemática de la relación funcional que representa las reglas de clasificación en términos de las variables medidas (o, alternativamente, en la forma geométrica de los límites de decisión en el espacio multidimensional). En este marco, la principal diferenciación se hace entre métodos lineales y no lineales. En los métodos lineales, las reglas de clasificación dan como resultado límites de decisión que son funciones lineales de las variables originales (es decir, que corresponden a

superficies lineales en el hiperespacio abarcado por las variables: una línea en dos dimensiones, un plano en tres dimensiones y un hiperplano en más de tres dimensiones). Por otro lado, las reglas de clasificación calculadas por los métodos no lineales dan como resultado hipersuperficies más complejas que separan las regiones correspondientes a las diferentes categorías en el espacio multidimensional de las variables, por ejemplo, en el caso de los métodos cuadráticos, los límites de decisión pueden ser hiperelipsoides, hiperparaboloide o hiperesferas [10].

Existe un gran número de algoritmos supervisados desarrollados bien por la estadística (Regresión Logística, Análisis Discriminante) o bien por la inteligencia artificial (Redes Neuronales, Inducción de Reglas, Árboles de Decisión, Redes Bayesianas) capaces de realizar las tareas propias de la clasificación. Sin embargo, a continuación se explicarán los métodos empleados en la realización de la presente tesis doctoral.

- *Análisis discriminante lineal (LDA)*

El Análisis Discriminante Lineal o Linear Discriminant Analysis (LDA), fue la primera técnica de clasificación multivariada, desarrollada por Fisher en 1936, convirtiéndose en el método de reconocimiento de patrones supervisado más antiguo y estudiado en la literatura. Como su nombre indica, es una técnica lineal, es decir, los límites de decisión que separan las clases en el espacio multidimensional de las variables son superficies lineales (hiperplanos). Desde un punto de vista probabilístico, es un método paramétrico, ya que su hipótesis subyacente es que, para cada categoría, los datos siguen una distribución normal multivariada [4,10].

Existen dos enfoques en la clasificación discriminante:

- el basado en la obtención de funciones discriminantes de cálculo similar a las ecuaciones de regresión lineal múltiple.
- empleando técnicas de correlación canónica y de componentes principales, denominado análisis discriminante canónico.

El primer enfoque es el más comúnmente empleado y es el que abordaremos en este punto. En este método, el modelo se construye utilizando los datos de los objetos precategorizados en categorías conocidas (conjunto de datos de entrenamiento), para encontrar una combinación lineal para discriminar clases de objetos [23].

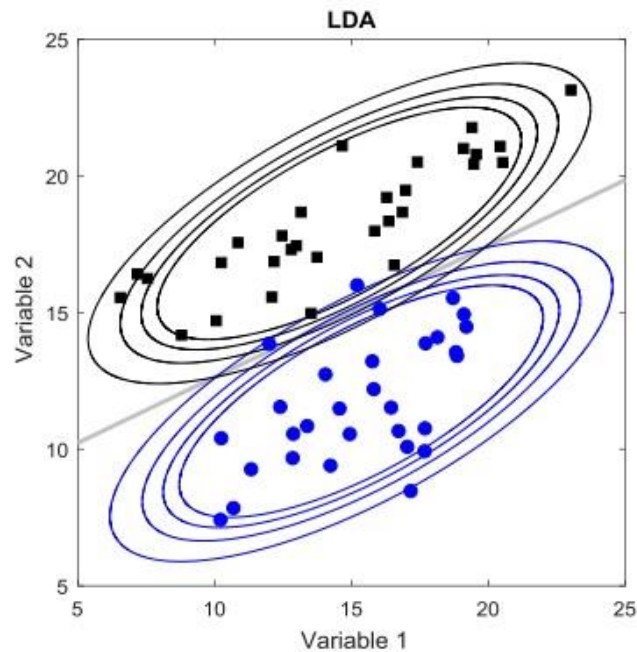


Figura 1.20. Delimitadores para un caso bivariado simulado de LDA.

Su fundamento matemático está en conseguir, a partir de las variables explicativas o independientes, unas funciones lineales de éstas con capacidad para clasificar a otros individuos, donde la función de mayor valor define el grupo al que pertenece de forma más probable.

LDA se basa en la descripción de datos mediante distribuciones de densidad de probabilidad, bajo la hipótesis de que son multivariantes normales y con la misma dispersión y correlación entre variables dentro de todas las clases. Dichas funciones de densidad de probabilidad se estiman a partir de datos experimentales utilizando parámetros estadísticos descriptivos que definen su posición y su dispersión: a saber, media/centroides como parámetros de ubicación, varianzas y covarianzas como parámetros de dispersión. En el caso de dos variables, la función de densidad de probabilidad tiene forma de campana y sus líneas

de sección elíptica, como se presenta en la Figura 1.20, corresponden a valores de densidad de probabilidad iguales. Extendiendo dicho método a sistemas multivariados, la región de isoprobabilidad se convierte en un hiper-elipsoide. Las elipses de diferentes categorías poseen excentricidades y orientaciones de eje iguales, debido a las hipótesis detrás de LDA. Al conectar los puntos de intersección de cada par de elipses correspondientes (en el mismo nivel de confianza), se identifica una línea recta (Figura 1.20): esta línea representa el delimitador entre las dos clases [24].

- *Análisis discriminante de mínimos cuadrados parciales (PLS-DA)*

El Análisis Discriminante de Mínimos Cuadrados Parciales o Partial Least Squares Discriminant Analysis (PLS-DA) es una herramienta de clasificación lineal que calcula modelos predictivos basados en el algoritmo de regresión de mínimos cuadrados parciales que busca variables latentes con máxima covarianza y representa las fuentes relevantes de variabilidad de datos con combinaciones lineales de las variables originales [12,25].

El método fue introducido por Barker y Rayens en 2003 y aplica la regresión de mínimos cuadrados parciales (PLS) utilizando como variable de respuesta un índice ficticio o dummy que asocia el mismo número entero a muestras que pertenecen a la misma clase [25]. Cuando hay más de dos clases involucradas, se aplica el algoritmo PLS-2; este algoritmo permite predecir una matriz de variables de respuesta, una para cada clase, simultáneamente. El método PLS-DA proporciona un delimitador lineal (Figura 1.21) que se ha demostrado que es análogo al obtenido por LDA [18].

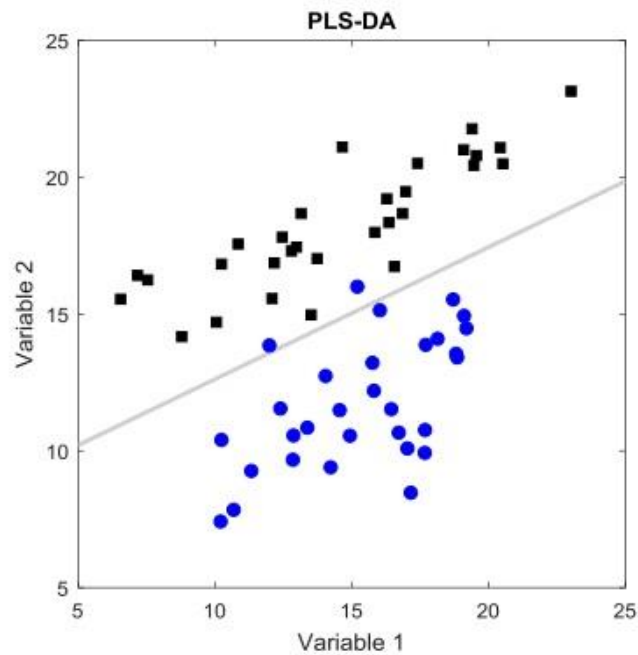


Figura 1.21. Gráfico de un caso simulado de PLS-DA.

- *Regresión logística (LR)*

La Regresión Logística o Logistic Regression (LR) se usa para modelar la probabilidad de ocurrencia de un evento (una variable binaria, como sí/no o 1/0), utilizando predictores categóricos o numéricos, es decir, relaciones entre variables independientes y una variable dependiente que solo puede tomar un valor entero y finito de clases. En términos generales, la regresión logística calcula las probabilidades de ocurrencia de alguna de las clases del modelo a partir del uso de la función logística [17].

La Regresión Logística Multinomial (MLR) es utilizada en modelos con la variable dependiente de tipo nominal con más de dos categorías (politómica) y es una extensión multivariante de la regresión logística binaria clásica. MLR basado en el modelo logístico, pertenece a la familia de modelos lineales generalizados y se utiliza cuando la variable de respuesta es una variable categórica. Es aplicable para conjuntos de datos en los que la variable dependiente puede tener más de dos opciones que están codificadas categóricamente, una de las categorías se toma como la categoría de referencia y cuando las variables independientes son predictores continuos o categóricos. MLR predice la probabilidad de

ocurrencia de un evento ajustando los datos a una función logit y asegura que todas las probabilidades estén en el intervalo [0, 1] y que las probabilidades sumen 1. El modelo puede describir las relaciones entre un conjunto de variables predictoras y una variable dependiente dicotómica o multinomial [5,26]. Con el propósito de transformar un conjunto de modelos de regresión logística binomial en un modelo multinomial generalizado, se utiliza la siguiente ecuación:

$$p_i = \frac{\exp(p_i)}{\exp(p_1) + \exp(p_2) + \dots + \exp(p_n)}$$

donde p_i representa la probabilidad de ocurrencia para la clase i , y el denominador de la ecuación representa la suma de las probabilidades de ocurrencia para n clases, por lo tanto, en la clasificación final, cada punto de datos se envía a la clase con la mayor probabilidad de ocurrencia [5,26,27].

- *K-Vecinos más cercanos (K-NN)*

El método de K-Vecinos más cercanos o K-Nearest Neighbors (K-NN) es un método no paramétrico basado en la distancia, utilizado para fines de clasificación. Se utiliza para mejorar el análisis discriminante cuando las estimaciones paramétricas confiables de las densidades de probabilidad son desconocidas o difíciles de determinar y, para clasificar una muestra desconocida en función de su proximidad a las muestras ya ubicadas en categorías utilizando la distancia euclidiana entre ellas. Por lo tanto, las predicciones se hacen en función de un vecindario que se define por el número k de puntos de entrenamiento que se encuentran más cerca del punto predicho. Para las variables de respuesta continua, la predicción de vecinos más cercanos, \tilde{y}_i , para el conjunto de objetivos i -ésimo cuyas predicciones de variables de respuesta se desean se calcula como:

$$\tilde{y}_i = \sum_{j=1}^k w_{ij} y_j^i$$

donde $\{y_j^i, j = 1, 2, \dots, k\}$ es el conjunto de observaciones de variables de respuesta para los elementos del conjunto de referencia k más cercanos o más similares al i -ésimo elemento del

conjunto de objetivos en variables de características con respecto a una métrica de distancia , y w_{ij} es el peso asignado a los j -ésimos vecinos más cercanos. Para las variables categóricas, la clase predicha del i -ésimo elemento objetivo es la clase más ponderada entre los k vecinos más cercanos, una mediana ponderada o un modo en el caso de variables de escala ordinal y un modo en el caso de variables nominales [28].

Resumiendo, el algoritmo de vecino más cercano se basa en la determinación de las distancias entre un objeto desconocido y cada uno de los objetos del conjunto de entrenamiento. Por lo general, se utiliza la distancia euclidiana, pero para las variables fuertemente correlacionadas, se prefieren las medidas basadas en la correlación. Luego, se selecciona la distancia más corta para la asignación de la membresía de la clase. La elección de k se optimiza calculando la capacidad de predicción con diferentes valores de k . Se prefieren los valores de k pequeños (3 o 5). Se requiere un pre-procesamiento de los datos para evitar el efecto de diferentes escalas de las variables. La técnica también se puede aplicar a las puntuaciones de las muestras relacionadas con los componentes principales.

El método presenta varias ventajas: su simplicidad matemática, que no impide que logre resultados de clasificación tan buenos (o incluso mejores) que otras técnicas de reconocimiento de patrones más complejas; está libre de supuestos estadísticos, como la distribución normal de las variables; y su efectividad no depende de la distribución espacial de las clases [29].

- *Redes neuronales (NNET)*

Las redes neuronales o neural networks (NNET) se definen como estructuras compuestas de elementos de procesamiento adaptativos simples densamente interconectados, llamados neuronas (o nodos) artificiales que son capaces de realizar cálculos masivamente y en paralelo para el procesamiento de datos y la representación del conocimiento. NNET incluye una gran clase de métodos de aprendizaje que se desarrollaron por separado en estadística e inteligencia artificial. El método más utilizado es el perceptrón de capa única, también red de

propagación posterior de capa oculta única. Detrás de este nombre complicado y con terminología relacionada con la biología, NNET es sólo un modelo estadístico no lineal [14,30].

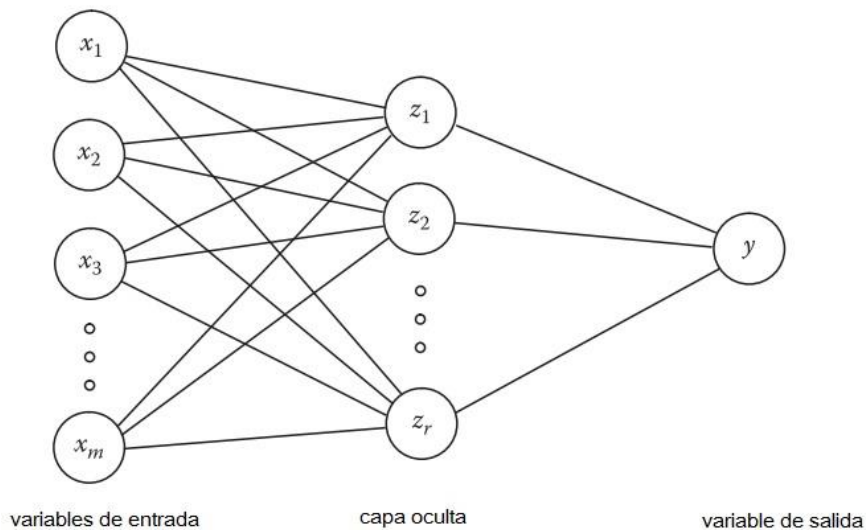


Figura 1.22. Esquema de una red neuronal con una sola capa oculta.

NNET es un modelo computacional y matemático inspirado en el sistema nervioso biológico. Por lo tanto, algunos de los términos utilizados en una red neuronal artificial se toman prestados de sus homólogos biológicos. Las redes neuronales artificiales de retroalimentación se componen de una gran cantidad de neuronas artificiales dispuestas en una estructura paralela (Fig. 1.22). Las neuronas se clasifican en una capa de entrada, que contiene una neurona para cada variable independiente (x_m); una o más capas ocultas, donde se procesan los datos; y una capa de salida, con una neurona para cada variable dependiente (y). Por lo tanto, los datos de la capa de entrada se propagan a través de la red vía sinapsis, que están asociadas con coeficientes de conectividad llamados pesos (w). La entrada se calcula como el producto interno de las variables de entrada (x_m) que inciden en la neurona y sus pesos (w). Una vez que el nodo calcula este producto, el resultado (z_r) se pasa a una función de transferencia (activación) no lineal diferenciable, que transforma la suma ponderada de todas las variables que inciden en la neurona para obtener la salida valor (y). Se han utilizado varias funciones de transferencia, por ejemplo, funciones logísticas variantes, tangente hiperbólica o

función sigmoidea. El proceso de aprendizaje identifica los pesos que producen el mejor ajuste de los resultados predichos en todo el conjunto de datos de entrenamiento. Las capas ocultas son particularmente importantes para hacer frente a problemas de clasificación no lineal [29].

En la construcción de una red neuronal se debe prestar especial atención a la selección de la arquitectura, la selección de los parámetros de aprendizaje y la validación de la red. Durante el paso de entrenamiento, se calculan los valores para todos los parámetros involucrados en el proceso de aprendizaje, el error entre la salida neta prevista y la salida correcta. El tamaño y el número de las capas ocultas y el número ciclos de entrenamiento se evalúan probando diferentes valores (prueba y error) y verificando la precisión de la predicción resultante. El número óptimo de ciclos para un NNET se logra cuando el error en el conjunto de prueba alcanza un mínimo. El número de nodos ocultos es crítico para el diseño de la red, porque si se usan demasiados nodos ocultos, la red se sobreajustará o memorizará los datos del conjunto de entrenamiento (ruido). Por el contrario, si se utilizan pocos nodos ocultos, la red no podrá generalizarse y se volverá inestable. Generalmente, una capa oculta es suficiente para aproximar funciones continuas mientras que dos capas ocultas pueden ser necesarias para aprender funciones con discontinuidades [29].

- *Máquina de vectores de soporte (SVM)*

La Máquina de Vectores de Soporte o Support Vector Machine (SVM) se refiere a una técnica de aprendizaje automático que puede ser usada para resolver problemas de clasificación y regresión lineales y no lineales utilizando la misma metodología. El clasificador SVM construye un hiperplano de separación óptimo, en el espacio de características o variables, entre varias clases [31]. En el contexto de la clasificación, producen límites lineales entre grupos de objetos en un espacio transformado de las variables x , que generalmente es una dimensión más alta que el espacio x original. La idea es hacer que los grupos sean linealmente separables, donde SVM detecta los puntos más cercanos entre dos clases en el espacio de características y asigna un margen basado en la distancia entre el hiperplano y los

puntos. Además, los límites de clase se construyen para maximizar el margen entre grupos mediante "vectores de soporte" y reducir las clases superpuestas. Este hiperplano, desde los márgenes maximizados, se usa como criterio para una clasificación posterior [5,32]. Los vectores de soporte son los puntos de cada clase que están más cerca al hiperplano de máximo margen; cada clase debe tener al menos un vector de soporte, pero es posible que exista más de uno. Utilizando estos vectores de soporte es posible definir el hiperplano de máximo margen (Fig. 1.23).

Una máquina de soporte vectorial puede ser entendida como una superficie que crea un límite entre los puntos de los datos graficados en un espacio multidimensional que representan ejemplos y los valores de sus atributos. El objetivo de la técnica SVM es crear un límite llamado hiperplano, que divide el espacio para crear divisiones apenas homogéneas en cada lado. De esta forma, el aprendizaje mediante SVMs combina aspectos del aprendizaje del tipo k-N vecinos más cercanos (k-NN) y de regresión lineal. La combinación es extremadamente poderosa, permitiendo al algoritmo SVM modelar relaciones altamente complejas.

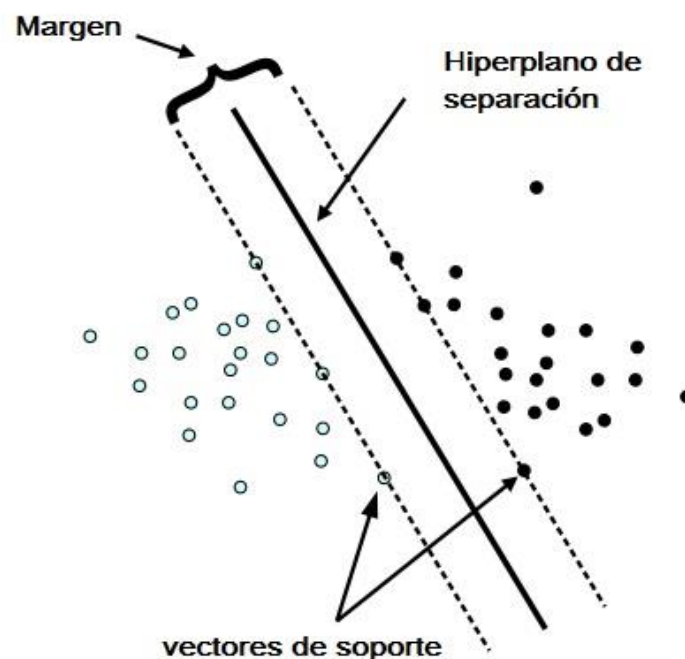


Figura 1.23. Esquema de los vectores de soporte.

SVM encuentra el hiperplano óptimo utilizando las muestras de entrenamiento e_1, e_2, \dots, e_n de cada clase que, en el espacio n -dimensional, tienen lugar más cerca del límite de decisión. Este límite de decisión viene dado por la siguiente ecuación:

$$w \cdot x + b = 0$$

donde x es el grupo de variables de las muestras de entrenamiento y w y b son parámetros del modelo. El margen del límite de decisión se calcula mediante la ecuación:

$$d = \frac{2}{\|w\|}$$

En SVM, los problemas no lineales se transforman en lineales al mapear los datos en un espacio de características de alta dimensión a través de una función lineal (función kernel) o truco del kernel y realizar una clasificación lineal o regresión en este espacio de características. Para problemas de clasificación, la dimensionalidad del espacio de características está determinada por la elección de la función kernel y su parámetro (γ), mientras que la complejidad del modelo está determinada por un parámetro de penalización adicional (C) [19]. Cuando el conjunto de datos no es separable mediante un hiperplano, se hace necesario relajar las restricciones de la ecuación de optimización. Esto se logra agregando un término más en la ecuación de optimización, con ϵ_i junto a el parámetro de coste, C . Un valor de coste, denotado por C , se aplica a todos los puntos que violan las restricciones impuestas y en vez de buscar el máximo margen de separación, el algoritmo intenta minimizar el valor de coste, C .

- *Bosques aleatorios (RF)*

El método de Bosques aleatorios o Random Forest (RF) es un algoritmo de aprendizaje automático utilizado para varios tipos de tareas de clasificación. RF consiste en una combinación de clasificadores estructurados en árbol, donde cada clasificador se genera usando un vector aleatorio muestreado independientemente del vector de entrada, y cada árbol del bosque emite un voto de unidad, asignando cada entrada a la etiqueta de clase más

probable para clasificarlo [33]. En otras palabras, árboles sucesivos que no dependen de árboles anteriores se construyen de forma independiente utilizando una muestra de arranque del conjunto de datos donde cada nodo se divide utilizando el mejor subconjunto de predictores elegidos aleatoriamente en ese nodo y, al final, se toma un voto de mayoría simple para la predicción, es decir una combinación de árboles de decisión que se construye usando diferentes conjuntos de variables seleccionadas aleatoriamente [34]. Dentro de las ventajas de este método, se puede mencionar la velocidad de los procesos de entrenamiento y predicción, es robusto al ruido, es un método exitoso que puede identificar patrones no lineales en los datos y puede manejar fácilmente datos numéricos y categóricos. Una de las principales ventajas de Random Forest es que no sufre sobreajuste, incluso si se añaden más árboles al bosque [35]. Otra característica relevante de este modelo es que también puede emplearse para fines de selección de las variables, ya que toma medidas implícitas de importancia de las variables (RFI). RF estima la importancia de una variable al observar cuánto aumenta el error de predicción cuando los datos para esa variable se permutan mientras que todos los demás se mantienen sin cambios. Los cálculos requeridos se realizan árbol por árbol a medida que se construye el bosque aleatorio [34].

La Fig. 1.24 presenta gráficamente el funcionamiento del método de RF: seis árboles de decisión que forman un bosque aleatorio (muy pequeño) para la clasificación; los árboles A, B y E se asignan a la clase roja, sin embargo, los árboles C y D se asignan a la clase verde y el árbol F se asigna a la clase amarilla, de modo que el bosque aleatorio clasificará el objeto como rojo por mayoría.

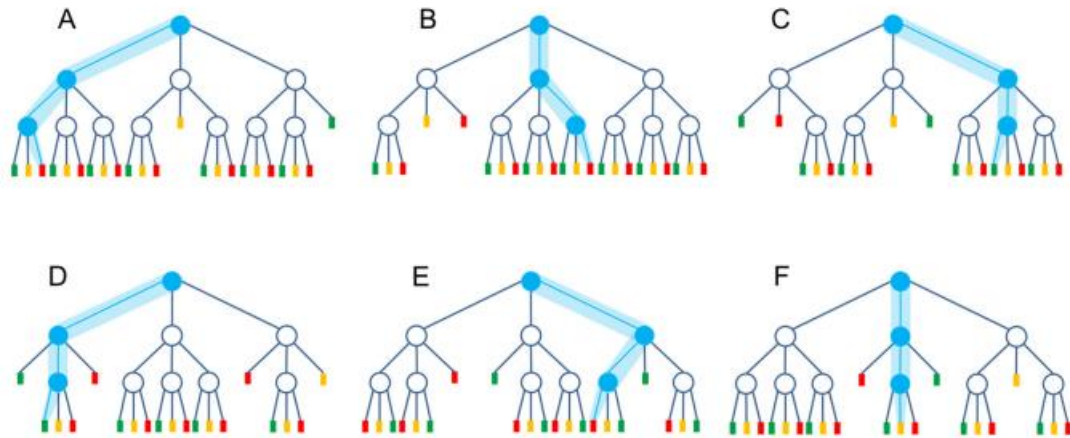


Figura 1.24. Descripción gráfica de cómo RF realiza la clasificación [6].

En resumen, el modelo de bosques aleatorios es una buena elección por varios motivos. A menudo, muy poco pre-procesamiento de los datos es necesario realizar, los datos no necesitan ser normalizados y es bastante robusta a los datos extremos. La necesidad de selección de variables se evita ya que el algoritmo lo realiza durante el análisis. Como muchos árboles se construyen a dos niveles de aleatoriedad (observaciones y variables), cada árbol es un modelo independiente y el modelo resultante tiende a no sobreajustar el set de datos de entrenamiento.

1.4.3.3. Validación del modelo

1.4.3.3.1. Estrategias de validación

La validación es una etapa clave del modelado quimiométrico clasificatorio, ya que es el conjunto de operaciones que permite evaluar la calidad de un modelo y cuán confiables es a la hora de clasificar nuevas muestras. La validación de los modelos obtenidos por técnicas supervisadas es uno de los aspectos más importantes del reconocimiento de patrones. Esta etapa implica la evaluación del número de variables o componentes significativos necesarios para caracterizar el conjunto de datos, la capacidad de predicción del modelo para muestras desconocidas y el carácter representativo de los datos utilizados para producir el modelo [29].

La utilidad de un modelo de clasificación siempre debe estar determinada por su confiabilidad en la predicción. La estimación de la capacidad de predicción en nuevos objetos,

es decir, muestras no utilizadas para construir los modelos, es un paso fundamental en cualquier proceso de modelado, y se han implementado varios procedimientos para este propósito. Las estrategias de validación más comunes dividen los objetos disponibles en dos subconjuntos: un conjunto de entrenamiento (o calibración) utilizado para calcular el modelo y un conjunto de evaluación externo utilizado para evaluar su confiabilidad (testeo). Una característica clave para una validación honesta es que los objetos en el conjunto de evaluación externa tienen que ser absolutamente ajenos al modelo: no se puede usar información de ellos para construir el modelo; de lo contrario, la capacidad de predicción puede sobreestimarse [4]. Además, en la fase de validación, los datos de validación proporcionan una evaluación imparcial del modelo entrenado mientras ajustan los hiperparámetros del modelo.

La situación ideal es cuando hay suficientes muestras disponibles para crear conjuntos de entrenamiento y prueba independientes y cada conjunto contiene muestras representativas de cada clase. Este procedimiento de validación se conoce como validación externa y como se expuso anteriormente, en este tipo de validación el conjunto de pruebas es completamente independiente del proceso de construcción del modelo (selección de variables, estimación de parámetros, determinación de componentes principales). Pero esta situación ideal no suele ser común; por lo tanto, los métodos de validación cruzada se usan comúnmente. La validación cruzada (CV) es probablemente el procedimiento de validación más utilizado. En la validación cruzada o cross-validation, la capacidad de predicción del modelo se determina desarrollando un modelo con parte del conjunto de datos (conjunto de entrenamiento o aprendizaje) y utilizando otra parte de datos (conjunto de prueba) para probando el modelo. Tanto el conjunto de entrenamiento como el de prueba contienen muestras representativas de cada clase. Este procedimiento, que consiste en el desarrollo del modelo y la prueba del modelo, se repite varias veces para que las mismas muestras tengan la

probabilidad de ser utilizadas como entrenamiento y como objetos de prueba. Uno de estos métodos es la validación cruzada de k-iteraciones o k-fold cross-validation (Fig. 1.25) [29].

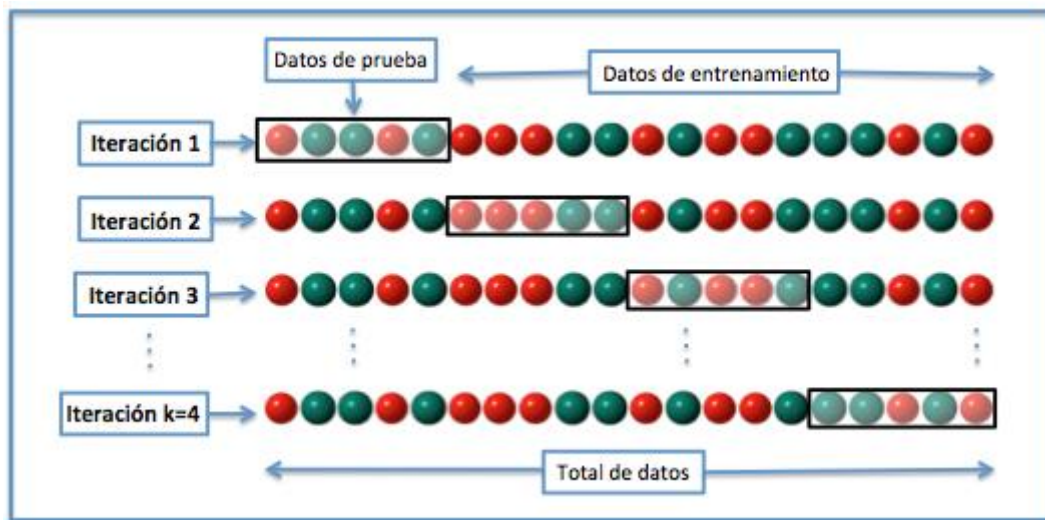


Figura 1.25. Validación cruzada de k= 4 iteraciones.

En la validación cruzada de K iteraciones los datos de muestra se dividen en K subconjuntos. Uno de los subconjuntos se utiliza como datos de prueba y el resto (K-1) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones, con diferentes constituciones de ambos conjuntos para garantizar que todas las muestras tengan la posibilidad de ser incluidas en el conjunto de evaluación al menos una vez. Finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado [29].

Una alternativa un poco menos popular, pero aún bastante ampliamente utilizada para k-fold CV se conoce como muestreo bootstrap, bootstrap o bootstrapping para abreviar. En términos generales, se refieren a métodos estadísticos de uso de muestras aleatorias de datos para estimar las propiedades de un conjunto más grande. Cuando este principio se aplica al rendimiento del modelo de aprendizaje automático, implica la creación de varios conjuntos de datos de prueba y entrenamiento seleccionados al azar, que luego se utilizan para estimar las estadísticas de rendimiento. Los resultados de los diversos conjuntos de datos aleatorios se promedian para obtener una estimación final del rendimiento futuro. El bootstrap permite

seleccionar ejemplos varias veces a través de un proceso de muestreo con reemplazo. Esto significa que a partir del conjunto de datos original de n ejemplos, el procedimiento bootstrap creará uno o más nuevos conjuntos de datos de entrenamiento que también contienen n ejemplos, algunos de los cuales se repiten. Los conjuntos de datos de prueba correspondientes se construyen a partir del conjunto de ejemplos que no se seleccionaron para los conjuntos de datos de entrenamiento respectivos.

1.4.3.3.2. Evaluación de los métodos quimiométricos

Para evaluar el rendimiento de los clasificadores discutidos se utilizan los siguientes parámetros:

- Matriz de confusión

Una matriz de confusión es una tabla de contingencia que clasifica las predicciones según si coinciden con el valor real de los datos. Según el número de clases, hay dos tipos de problemas de clasificación, la clasificación binaria donde solo hay dos clases y la clasificación de clases múltiples donde el número de clases es mayor que dos. Si existen dos clases, es decir, clasificación binaria, se expresa P para clase positiva y N para la clase negativa y entonces, una muestra desconocida se clasifica en P o N [36].

		Clase Real	
		Positivo (P)	Negativo (N)
Clase Predicha	Verdadero (V)	Verdadero Positivo (VP)	Falso Positivo (FP)
	Falso (F)	Falso Negativo (FN)	Verdadero negativo (VN)
		$P = VP + FN$	$N = FP + VN$

Figura 1.26. Ejemplo ilustrativo de una matriz de confusión.

En la Fig 1.26 se presenta la matriz de confusión para un caso binario que contiene cuatro valores característicos: verdaderos positivos (VP), falsos positivos (FP), falsos negativos

(FN), y verdaderos negativos (VN). Los casos verdaderos positivos (VP) y los verdaderos negativos (VN) son los casos que han sido clasificados correctamente mientras que los casos de falso negativo (FN) y falso positivo (FP) son casos clasificados erróneamente.

- *Exactitud*

La exactitud o accuracy es una de las medidas más utilizadas para evaluar el rendimiento de una clasificación, y se define como una relación entre las muestras correctamente clasificadas sobre el total de las muestras utilizadas en el proceso de clasificación.

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

- *Sensibilidad y Especificidad*

La sensibilidad, la tasa de verdaderos positivos (true positive rate, TPR) o tasa de aciertos de un clasificador representa las muestras positivas clasificadas correctamente para el número total de muestras positivas, y se estima de acuerdo con la siguiente ecuación:

$$\text{Sensibilidad} = \frac{VP}{VP + FN} = \frac{VP}{P}$$

La especificidad o tasa de verdaderos negativos (true negative rate, TNR) expresa la relación de las muestras negativas clasificadas correctamente al número total de muestras negativas como se presenta en la ecuación:

$$\text{Especificidad} = \frac{VN}{FP + VN} = \frac{VN}{N}$$

Por lo tanto, la especificidad representa la proporción de las muestras negativas que se clasificaron correctamente, y la sensibilidad es la proporción de las muestras positivas que se clasificaron correctamente.

La sensibilidad y la especificidad varían de 0 a 1, siendo más deseables valores cercanos a 1. Por supuesto, es importante encontrar un equilibrio apropiado entre los dos, una tarea que a menudo es bastante específica del contexto [36].

- *Precisión y Recall*

Estrechamente relacionadas con la sensibilidad y la especificidad hay otras dos medidas del rendimiento de la clasificación: precisión y recall. Utilizadas principalmente en el contexto de la recuperación de información, estas estadísticas están destinadas a proporcionar una indicación de cuán interesantes y relevantes son los resultados de un modelo, o si las predicciones se diluyen con ruido sin sentido.

La precisión (también conocida como valor de los positivos predichos, PPV) se define como la proporción de ejemplos positivos que son realmente positivos; en otras palabras, evalúa cuán preciso es el algoritmo para identificar los casos de una clase dada.

$$\text{Precisión} = \frac{VP}{VP + FP}$$

Por otro lado, recall es una medida de cuán completos son los resultados y se define como el número de verdaderos positivos sobre el número total de positivos. En esta definición se puede reconocer que es lo mismo que la sensibilidad, solo que la interpretación difiere.

- Valor Kappa

El estadístico kappa ajusta la exactitud teniendo en cuenta la posibilidad de una predicción correcta. Se puede definir como medir el grado de acuerdo entre dos conjuntos de datos categorizados. Los valores de Kappa varían a un valor máximo de 1, lo que indica un acuerdo perfecto entre las predicciones del modelo y los valores verdaderos, algo poco frecuente. Los valores inferiores a uno indican un acuerdo imperfecto [20].

Dependiendo de cómo se va a utilizar su modelo, la interpretación de la estadística kappa puede variar. Una interpretación común se muestra a continuación:

- Mal acuerdo = Menos de 0,20
- Acuerdo justo = 0,20 a 0,40
- Acuerdo moderado = 0,40 a 0,60
- Buen acuerdo = 0,60 a 0,80
- Muy buen acuerdo = 0,80 a 1,00

La siguiente es la fórmula para calcular la estadística kappa. En esta fórmula, Pr se refiere a la proporción actual (a) y esperado (e) de acuerdo entre el clasificador y los valores verdaderos:

$$k = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$

1.4.4. Referencias

- [1] N. Dordevic, Chemometrics and stable isotope ratios of wine, 2015. <http://repository.uhn.ru.nl/handle/2066/147420>.
- [2] M. Kamal, R. Karoui, Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review, Trends Food Sci. Technol. 46 (2015) 1–22. doi:10.1016/j.tifs.2015.07.007.
- [3] M. Forina, M. Casale, P. Oliveri, Application of Chemometrics to Food Chemistry, in: Compr. Chemom., Elsevier, 2009: pp. 75–128. doi:10.1016/B978-044452701-1.00124-1.
- [4] P. Oliveri, G. Downey, Chapter 13 - Discriminant and Class-Modelling Chemometric Techniques for Food PDO Verification, in: M. de la Guardia, A. González (Eds.), Compr. Anal. Chem., Elsevier, 2013: pp. 317–338. doi:<https://doi.org/10.1016/B978-0-444-59562-1.00013-X>.
- [5] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 2016.
- [6] A.M. Jiménez-Carvelo, A. González-Casado, M.G. Bagur-González, L. Cuadros-Rodríguez, Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – A review, Food Res. Int. 122 (2019) 25–39. doi:10.1016/j.foodres.2019.03.063.
- [7] C.M. Cuadras, Nuevos métodos de análisis multivariante, CMC, Barcelona, España, 2014.
- [8] M. Balzarini, C. Bruno, M. Córdoba, I. Teich, Herramientas en el Análisis Estadístico Multivariado, Escuela Virtual Internacional CAVILA. Facultad de Ciencias Agropecuarias,

- Universidad Nacional de Córdoba, Córdoba, Argentina, 2015.
- [9] J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, Third edit, Elsevier Inc., USA, 2012.
- [10] F. Marini, *Chemometrics in Food Chemistry*, First, Elsevier B.V., Amsterdam, 2013.
- [11] R.G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, First, John Wiley & Sons, Ltd., Chichester, England, 2003.
- [12] D. Ballabio, V. Consonni, *Classification tools in chemistry. Part 1: linear models. PLS-DA*, *Anal. Methods*. 5 (2013) 3790–3798. doi:10.1039/C3AY40582F.
- [13] D.B.R. Todeschini, *Multivariate Classification for Qualitative Analysis*, in: D.-W. Sun (Ed.), *Infrared Spectrosc. Food Qual. Anal. Control*, 1st Editio, Academic Press, 2009: pp. 83–102.
- [14] K. Varmuza, P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press, 2009.
- [15] A. González, M. de la Guardia, *Basic Chemometric Tools*, in: *Compr. Anal. Chem.*, 2013: pp. 299–315. doi:https://doi.org/10.1016/B978-0-444-59562-1.00012-8.
- [16] L. Mutihac, R. Mutihac, *Mining in chemometrics*, *Anal. Chim. Acta*. 612 (2008) 1–18. doi:10.1016/j.aca.2008.02.025.
- [17] R. Nisbet, G. Miner, K. Yale, *Classification*, in: *Handb. Stat. Anal. Data Min. Appl.*, Elsevier, 2018: pp. 169–186. doi:10.1016/B978-0-12-416632-5.00009-8.
- [18] R.G. Brereton, *Pattern recognition in chemometrics*, *Chemom. Intell. Lab. Syst.* 149 (2015) 90–96. doi:10.1016/j.chemolab.2015.06.012.
- [19] E. Bona, P.H. Março, P. Valderrama, *Chemometrics Applied to Food Control*, in: *Food Control Biosecurity*, Elsevier Inc., 2018: pp. 105–133. doi:10.1016/B978-0-12-811445-2.00004-0.
- [20] Y. Kumar, G. Sahoo, *Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA*, *Int. J. Inf. Technol. Comput. Sci.* 4 (2012) 43–49.

- doi:10.5815/ijitcs.2012.07.06.
- [21] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*, 2017.
- [22] G.M. Downs, J.M. Barnard, *Clustering Methods and Their Uses in Computational Chemistry*, in: *Rev. Comput. Chem. Vol. 18*, John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2003: pp. 1–40. doi:10.1002/0471433519.ch1.
- [23] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of chemometrics and qualimetrics, Vol.B*, Amsterdam: Elsevier, 1997.
- [24] P. Oliveri, C. Malegori, M. Casale, *Multivariate Classification Techniques*, in: *Ref. Modul. Chem. Mol. Sci. Chem. Eng.*, Elsevier, 2018: pp. 481–486. doi:10.1016/B978-0-12-409547-2.14239-8.
- [25] M. Barker, W. Rayens, *Partial least squares for discrimination*, *J. Chemom.* 17 (2003) 166–173. doi:10.1002/cem.785.
- [26] B. Heung, H.C. Ho, J. Zhang, A. Knudby, C.E. Bulmer, M.G. Schmidt, *An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping*, *Geoderma*. 265 (2016) 62–77. doi:10.1016/j.geoderma.2015.11.014.
- [27] B. Kempen, D.J. Brus, G.B.M. Heuvelink, J.J. Stoorvogel, *Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach*, *Geoderma*. 151 (2009) 311–326. doi:10.1016/j.geoderma.2009.04.023.
- [28] R.E. McRoberts, *Forest Ecology and Management Estimating forest attribute parameters for small areas using nearest neighbors techniques*, *For. Ecol. Manage.* 272 (2012) 3–12. doi:10.1016/j.foreco.2011.06.039.
- [29] L.A. Berrueta, R.M. Alonso-Salces, K. Héberger, *Supervised pattern recognition in food analysis*, *J. Chromatogr. A*. 1158 (2007) 196–214. doi:10.1016/j.chroma.2007.05.024.
- [30] O.A. Moldes, J.C. Mejuto, R. Rial-Otero, J. Simal-Gandara, *A critical review on the*

- applications of artificial neural networks in winemaking technology, *Crit. Rev. Food Sci. Nutr.* 57 (2017) 2896–2908. doi:10.1080/10408398.2015.1078277.
- [31] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer New York, 2009. doi:10.1111/j.1751-5823.2009.00095_18.x.
- [32] B. Heung, H.C. Ho, J. Zhang, A. Knudby, C.E. Bulmer, M.G. Schmidt, An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping, *Geoderma*. 265 (2016) 62–77. doi:10.1016/j.geoderma.2015.11.014.
- [33] L. Breiman, Random Forest, *Mach. Learn.* 45 (2001) 5–32. doi:10.1023/A:1010933404324.
- [34] a Liaw, M. Wiener, Classification and Regression by randomForest, *R News*. 2 (2002) 18–22. doi:10.1177/154405910408300516.
- [35] A. Chaudhary, S. Kolhe, R. Kamal, An improved random forest classifier for multi-class classification, *Inf. Process. Agric.* 3 (2016) 215–222. doi:10.1016/j.inpa.2016.08.002.
- [36] A. Tharwat, Classification assessment methods, *Appl. Comput. Informatics*. (2018). doi:10.1016/j.aci.2018.08.003.

CAPÍTULO II: OBJETIVOS E HIPÓTESIS

2.1. OBJETIVOS GENERALES

En el presente trabajo se propone identificar la presencia de marcadores químicos de trazabilidad en uvas y productos derivados, mediante técnicas de "huella dactilar" y quimiométricas, con especial énfasis en la determinación de la composición química multielemental, asistidas por métodos de análisis multivariado modernos para establecer modelos tendientes a resolver problemas de autenticación de uvas y productos derivados.

2.2. OBJETIVOS ESPECÍFICOS

- Desarrollar y aplicar métodos analíticos sensibles y selectivos para la determinación multielemental, de modo de explorar la presencia de marcadores químicos de identidad en uvas y productos derivados (ej. pasas, jugo, semilla y aceite de pepita) producidos en la región de Cuyo.
- Proponer modelos quimiométricos que permitan establecer la trazabilidad de las distintas muestras en base al empleo de las llamadas técnicas de huella dactilar multielemental.
- Generar información nutricional y de la calidad de los productos alimenticios mencionados en base al conocimiento de su composición química inorgánica (aspectos cuali- y cuantitativos de la presencia de elementos esenciales y tóxicos).
- Proponer modelos matemáticos útiles para la industria vitícola regional que permitan certificar autenticidad, origen, posibles adulteraciones y/o varietal de futuras muestras.

2.3. HIPÓTESIS

La combinación de técnicas analíticas de huella dactilar multielemental con técnicas de análisis multivariado permite generar modelos tendientes a establecer la trazabilidad y denominación de origen de agroalimentos y sus productos derivados.

CAPÍTULO III:
REACTIVOS, MATERIALES
Y MÉTODOS

3.1. REACTIVOS, MATERIALES E INSTRUMENTACIÓN

3.1.1. Reactivos y materiales

3.1.1.1. Agua ultrapura

En el desarrollo experimental de la presente tesis doctoral se utilizaron dos sistemas de purificación de agua. El agua ultrapura fue obtenida mediante los equipos de purificación de agua Milli-Q (Millipore, París, Francia) y OSMOION (APEMA, Buenos Aires, Argentina). Se utilizó agua bidestilada para la carga de los purificadores, lo cual aseguró la calidad del agua ultrapura obtenida con un valor de resistividad de 18 M Ω -cm.

3.1.1.2. Soluciones patrón y estándares internos para mediciones en ICP-MS

Las diluciones de las soluciones patrones para los pasos de calibración y control de las mediciones y los estándares internos utilizados para corregir derivas instrumentales durante el análisis de las muestras por ICP-MS se prepararon a partir de:

- soluciones estándar multielemental certificadas 2, 3 y 5, y solución mono-elemental de rodio (^{103}Rh) como estándar interno de Perkin Elmer Pure Plus Atomic Spectroscopy Standards, (Norwalk, USA) para las mediciones en semillas y hollejos de uva.
- soluciones estándar multielemental certificadas (20 ppm de Al, Sb, As, Ba, Be, B, Cd, Ca, Cr, Co, Cu, Fe, Pb, Li, Mg, Mn, Mo, Ni, Se, Na, Sr, Tl, Sn, V y Zn, 100 ppm de P, K y Si y 5 ppm de Ag) y soluciones mono-elementales estándar de rodio (^{103}Rh) y holmio (^{165}Ho) como estándares internos de Chem-Lab, Ionex Reference Standard (Zedelgem, Bélgica) para los análisis de los jugos de uva.
- soluciones estándares multi-elementos CMS-4 (10 ppm de As, Be, Ga, Sb, V, B, Bi, In, Se, Ba, Cd, Pb y Tl), CMS-5 (10 ppm de Ag, Co, Cu, Li, Na, Sr, Al, Cr, Fe, Mg, Ni, Zn, Ca, Ce, K, Mn, Rb) y MSHG-100 ppm (Hg) de Inorganic Ventures (Christiansburg, USA) y soluciones mono-elementales de 1000 ppm de Sn, Mo, Pt y Pd de Centipur materiales de referencia de Merck (Darmstadt, Alemania). Se empleó una mezcla estándar interna para sistemas ICP-MS con 100 ppm de 6-Li, Sc, Ge, Rh, In, Tb, Lu, Bi de Agilent

Technologies (Santa Clara, USA). Todos estos estándares se utilizaron para los análisis de las muestras de vino.

3.1.1.3. Reactivos generales

- Ácido nítrico (65% (p/p), Merck, Darmstadt, Alemania)
- Peróxido de hidrógeno (30% (p/v), Biopack, Argentina)

Todos los reactivos que se utilizaron para el desarrollo experimental de la presente tesis doctoral fueron de calidad "reactivo analítico" o superior.

3.1.1.4. Material de vidrio

Se dispuso del material de vidrio necesario tanto volumétrico como de uso general. En todos los casos los materiales fueron lavados previamente a su uso mediante su inmersión durante 24 hs en una solución de HNO₃ al 10% (v/v) seguido de agua bidestilada y finalmente agua ultrapura.

3.1.2. Instrumentación

3.1.2.1. Espectrómetro de masas por plasma acoplado inductivamente

Para realizar las mediciones analíticas en los distintos tipos de muestras analizadas en la presente tesis doctoral, se utilizaron tres equipos de ICP-MS, los cuales se detallan a continuación:

- *Muestras de semillas y hollejos:*

Para la medición multielemental de las muestras de semillas y hollejos de uva se empleó un espectrómetro de masas de plasma acoplado inductivamente modelo PerkinElmer-SCIEX, ELAN DRC-e (Thornhill, Canadá) equipado con un nebulizador de perfluoracetato (PFA) resistente a HF y de alto rendimiento, acoplado a una cámara de pulverización ciclónica confeccionada con cuarzo, enfriada con el sistema PC³ de ESI (EE.UU.). Se utilizaron conos de muestreo y skimmer de níquel.

- *Muestras de jugo de uva:*

El análisis elemental de las muestras de jugo de uva digeridas se realizó en un modelo Nexlon 300X ICP-MS Perkin-Elmer, provisto de un muestreador automático CETAC AXS-520, conos de muestreo y skimmer de níquel y un nebulizador concéntrico.

- *Muestras de vinos:*

Las determinaciones analíticas en las muestra de vino se realizaron utilizando un instrumento ICP-MS modelo 7900 de Agilent (Tokio, Japón) equipado con un muestreador automático Agilent SPS 4, nebulizador, MicroMist, vidrio de borosilicato, conector de muestra UniFit, serie U , acoplado a una cámara de pulverización (cuarzo) y conos de muestreo y skimer de níquel. Todos los elementos fueron medidos con celda de colisión con gas Helio.

3.1.2.2. Plancha calefactora

Para el proceso de mineralización de los distintos tipos de muestra bajo estudio en la presente tesis doctoral, se empleó una plancha calefactora marca Thorbell (Santa Fe, Argentina), modelo FP 3030-T.

3.1.2.3. Medición de microvolúmenes

Las mediciones de microvolúmenes para la preparación de soluciones se realizaron con pipetas automáticas de succión y desplazamiento de aire (Eppendorf AG, Hamburgo, Alemania).

3.2. MUESTREO Y TRATAMIENTO DE LAS MUESTRAS

3.2.1. Hollejos y semillas de uva

Las muestras de uvas (*Vitis vinifera L.*) se recolectaron en viñedos ubicados en cinco regiones vinícolas de Mendoza (Argentina): Rivadavia (33°11'S, 68°28'O), San Martín (33°04'S, 68° 19'O), Guaymallén (32° 54'S, 68° 47'O), Junín (33° 15'S, 68° 43'O) y Maipú (32° 58'S, 68° 46'O), durante la cosecha de 2011. Las ubicaciones de los puntos de muestreo se muestran en la Fig. 3.1. Se recolectaron las siguientes variedades tintas: Cabernet Sauvignon, Malbec,

Bonarda, Aspirant Bouchet, y los siguientes varietales blancos: Chardonnay, Sauvignon Blanc y Pedro Jiménez. Para obtener muestras representativas, se recolectaron al azar racimos individuales de las parcela de los viñedos. Los racimos de uvas se cosecharon manualmente en su etapa de madurez óptima con un par de tijeras de podar apropiadas, cortándolas con cuidado en un punto del pedúnculo antes de llegar a la base del sarmiento (tallo de la vid). Posteriormente, las muestras se refrigeraron y se llevaron de inmediato al laboratorio para su análisis.

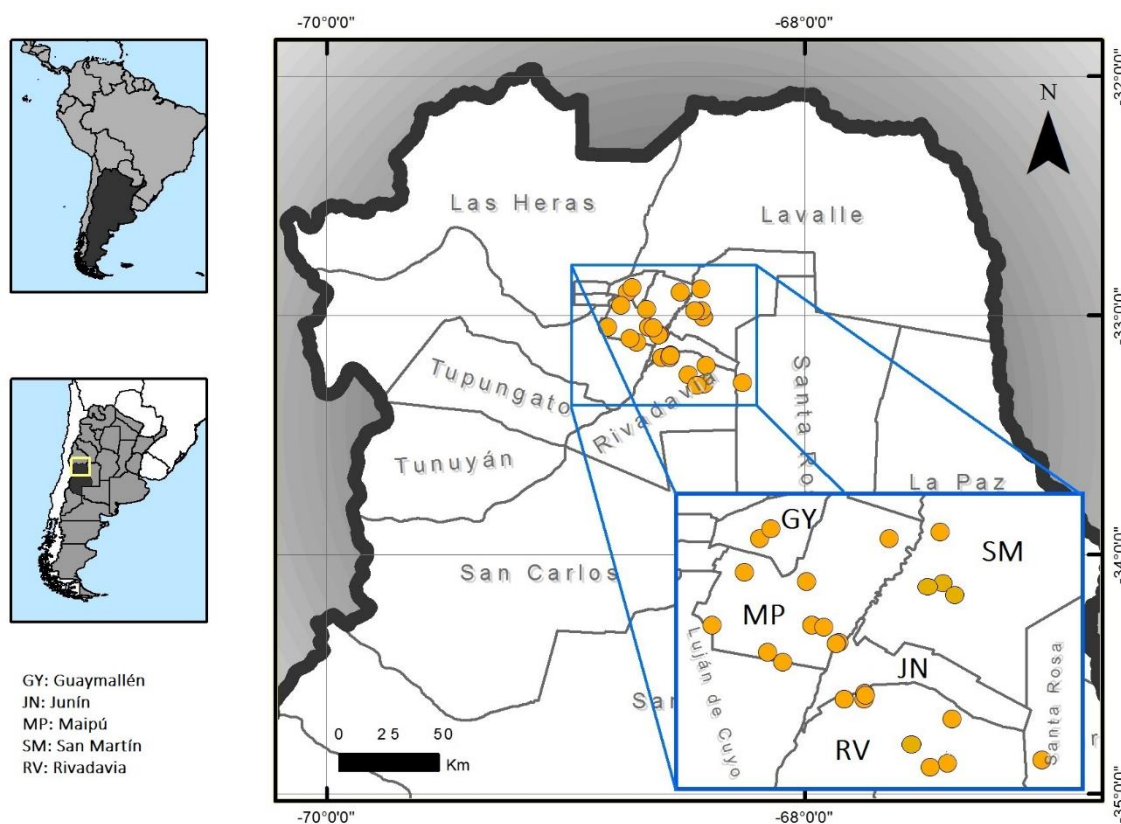


Fig. 3.1. Ubicación geográfica de los viñedos de muestreo seleccionados para la recolección de las muestras de uva.

Las uvas recolectadas se lavaron primero con agua potable y luego se enjuagaron con agua Milli-Q. La piel u hollejo y las semillas de las bayas de uvas se separaron manualmente, y luego se lavaron con agua Milli-Q para eliminar los residuos de pulpa. Posteriormente, las semillas y los hollejos se liofilizaron y finalmente se pulverizaron con un molino. Las muestras liofilizadas de cada parte de la baya se mineralizaron siguiendo el siguiente procedimiento: se

pesaron 0.5 g y se agregaron 6 mL de HNO₃ concentrado y 2 mL de H₂O₂ concentrado. Esta mezcla se dejó a temperatura ambiente durante 12 horas. Posteriormente, se aplicó el siguiente tratamiento térmico utilizando una plancha calefactora: 60 min a 50°C, 90 min a 100°C y finalmente 90 min a 150°C. Las muestras digeridas se dejaron enfriar hasta que se alcanzó la temperatura ambiente y luego se transfirieron a matraces volumétricos y se llevaron a un volumen final de 25 mL con agua Milli-Q. Posteriormente, se tomaron alícuotas de 2 mL de la solución anterior y se colocaron en un matraz de 50 mL y el volumen se completó con agua Milli-Q. Estas soluciones fueron analizadas por ICP-MS.

A continuación se presenta la Tabla 3.1 con los detalles de las muestras de hollejos y semillas finalmente analizadas.

Tabla 3.1. Detalles de las muestras de uva analizadas.

Departamento	Muestras		Ubicación	Varietal
	Hollejos	Semillas		
Rivadavia (RV)	90	92	La Central Rivadavia Los Campamentos Los Árboles Reducción	Sauvignon Blanc Chardonnay Aspirant Bouchet Cabernet Sauvignon Malbec
San Martín (SM)	74	69	Chapanay Tres Porteñas	Sauvignon Blanc Chardonnay Malbec Cabernet Sauvignon Bonarda
Guaymallén (GY)	22	25	Buena Nueva Colonia Segovia	Malbec
Junín (JN)	27	29	Medrano	Malbec Cabernet Sauvignon
Maipú (MP)	200	193	Barrancas Santa Blanca Isla Grande Lunlunta Fray Luis Beltrán Rodeo del Medio Cruz de Piedra Carrizal	Cabernet Sauvignon Malbec Bonarda Pedro Jiménez

3.2.2. Jugos de uva

Las muestras de jugos de uva de distintos varietales se adquirieron en diferentes mercados comerciales de Argentina (AR) y Brasil (BR), resultando un total de 26 muestras de las cuales 16 correspondieron a Argentina y 10 a Brasil. Las muestras se almacenaron a temperatura ambiente en los envases originales y antes del análisis se agitaron completamente para su homogenización. Para el análisis por ICP-MS, los jugos de uva se mineralizaron por digestión húmeda en sistema de recipiente abierto de la siguiente manera: una alícuota de 5 mL de cada muestra se puso en contacto con 2 mL de HNO₃ (65% p/p) y se dejó a temperatura ambiente por 30 min. Después de este paso, se aplicó un tratamiento térmico utilizando una plancha calefactora durante 45 minutos a 100°C. Posteriormente, las soluciones resultantes se dejaron enfriar y se transfirieron a matraces volumétricos completándose el volumen a 100 mL con agua ultrapura.

3.2.3. Vinos

Un total de 134 muestras de vinos se obtuvieron mediante colaboración con el Instituto Nacional de Tecnología Agropecuaria (INTA) sede Luján de Cuyo de Mendoza y el Instituto Nacional de Vitivinicultura (INV). Todas las muestras de vino correspondieron a varietales tintos y a diferentes años de cosecha (vintage). En la tabla 3.2 se detallan la procedencia y características generales de los vinos bajo estudio en la presente tesis doctoral.

Tabla 3.2. Detalles de los vinos analizados.

Región	Muestras	Procedencia	Varietales	Año
Mendoza – Centro-Oeste (Mza)	35	Luján de Cuyo	Cabernet Sauvignon	2010
		Lavalle	Malbec	2012
		Maipú	Bonarda	2013
		Las Heras	Cabernet Franc	2014
			Caladoc	2015
			Cereza	2016
			Mourvèdre	
			Syrah	
			Valency Merlot	

		Pinot noir Tempranillo		
Mendoza – Este (Mza)	14	San Martín	Ancellota	2010
		Junín	Caladoc	2012
		Santa Rosa	Malbec	2015
		Rivadavia	Nebbiolo	2016
			Bonarda	
Mendoza – Valle de Uco (Mza)	31	Tupungato	Malbec	2010
		San Carlos	Ancellota	2012
			Bonarda	2014
			Cabernet Franc	2015
			Cabernet Sauvignon	2016
			Caladoc	
			Corvina	
			Croatina	
			Grenache	
			Nebbiolo	
	Nero D'avola			
	Rondinella			
	Touriga Nacional			
Mendoza – Sur (Mza)	16	San Rafael	Ancellota	2012
		General Alvear	Bonarda	2013
			Cabernet Sauvignon	2015
			Corvina	2016
			Mourvèdre	2017
			Syrah	
Noroeste Argentino (Noa)	25	La Rioja (Valle de Famatina)	Malbec	2012
		Salta (Valle Calchaquíes y Valle de Cafayate)	Bonarda	2013
			Syrah	2014
		San Juan (Calingasta, Valle del Pedernal, Valle del Tulum, Valle del Zonda)	Merlot	2015
			Cabernet sauvignon	2016
			Tannat	2017
			Caladoc	2018
			Mourvèdre	
			Syrah	
Patagonia (Ptg)	13	La Pampa	Bonarda	2016
		Neuquén	Cabernet sauvignon	2017
		Río Negro	Malbec	2018
			Merlot	
			Pinot noir	
	Petit verdot			
			Tannat	

Las muestras se almacenaron a temperatura ambiente en los envases originales y antes del análisis se procedió a su homogenización. Para su análisis mediante ICP-MS, los vinos se mineralizaron mediante un tratamiento con ácido concentrado en vaso abierto. Para ello se

midió un volumen de 5 mL de las muestras, se dejaron por 24 hs a temperatura ambiente con 2,5 mL de HNO₃ (65% p/p), posteriormente se aplicó un tratamiento térmico mediante una plancha calefactora durante 45 minutos a 100°C. Finalmente las muestras digeridas se dejaron enfriar y se transfirieron a matraces volumétricos completándose el volumen final de 25 mL con agua ultrapura.

3.3. ANÁLISIS MULTIELEMENTAL MEDIANTE ICP-MS

Para el análisis multielemental se utilizó la técnica de espectrometría de masas con plasma acoplado inductivamente (ICP-MS). Se establece a continuación los detalles sobre el funcionamiento de cada equipo utilizado según el tipo de muestra.

3.3.1. Hollejos y semillas de uva

Las condiciones instrumentales utilizadas para el análisis de las semillas y hollejos de uva se presentan en la siguiente tabla:

Tabla 3.3. Condiciones óptimas instrumentales de la determinación multielemental en semillas y hollejos.

Instrumento	ELAN DRC-e (PerkinElmer SCIEX, Thornhill, Canada)
Flujo Ingreso muestra (mL min ⁻¹)	0,8
Introducción de la muestra	Nebulizador modelo PFA-ST, acoplado a cámara de nebulización: ciclónica, de cuarzo. Refrigerada a 4º con sistema PC3 de ESI (USA)
Potencia RF (W)	1000
Flujo del gas (L min ⁻¹)	Nebulizador: 0,85
Interfase	Conos de muestreo y skimmer de Ni
Adquisición de datos	Peak hopping
Dwell time (ms)	50
Número de réplicas	3
Isotopos monitoreados	⁴⁷ Ti, ⁵¹ V, ⁵⁵ Mn, ⁵⁷ Fe, ⁵⁹ Co, ⁶⁰ Ni, ⁶³ Cu, ⁶⁶ Zn, ⁶⁹ Ga, ⁷⁵ As, ⁸⁵ Rb, ⁸⁹ Y, ⁹⁰ Zr, ⁹³ Nb, ⁹⁸ Mo, ¹⁰⁷ Ag, ¹³⁰ Te, ¹³³ Cs, ¹³⁹ La, ¹⁴⁰ Ce, ¹⁴¹ Pr, ¹⁴² Nd, ¹⁵² Sm, ¹⁵³ Eu, ¹⁵⁸ Gd, ¹⁶⁹ Tm, ¹⁷⁵ Lu, ²⁰⁵ Tl y ²³⁸ U

El instrumento fue calibrado contra soluciones estándar externas certificadas. Para corregir las interferencias no espectrales, se agregó a cada muestra una solución de Rh (10 ppm) que se utilizó como estándar interno para todas las determinaciones. Se confeccionaron

rectas de calibración para cada elemento a partir de patrones por triplicado y empleando 5 niveles de concentración.

3.3.2. Jugos de uva

El análisis multielemental de las muestras de jugos de uva se llevó a cabo bajo las siguientes condiciones instrumentales del equipo de ICP-MS:

Tabla 3.4. Condiciones óptimas instrumentales de la determinación multielemental en jugos de uva.

Instrumento	PerkinElmer Nexlon 300X ICP-MS con CETAC ASX-520 Autosampler
Flujo Ingreso muestra (mL min ⁻¹)	0,85
Introducción de la muestra	MEINHARD® Concentric Nebulizer
Potencia RF (W)	1250
Flujo del gas (L min ⁻¹)	Nebulizador: 0,95; Plasma: 16; auxiliar: 1,2
Interfase	Conos de muestreo y skimmer de Ni
Adquisición de datos	Time resolved análisis
Dwell time (ms)	50
Número de réplicas	3
Isotopos monitoreados	⁵¹ V, ⁵² Cr, ⁵⁵ Mn, ⁵⁷ Fe, ⁶⁰ Ni, ⁶³ Cu, ⁷⁵ As, ⁸⁵ Rb, ⁸⁹ Y, ⁹⁸ Mo y ¹⁰⁶ Pd

Para la calibración del instrumento, se prepararon estándares de 0,5 hasta 100 ppb que luego se separaron en dos curvas de calibración (bajas concentraciones hasta 20 ppb y altas concentraciones de 20 a 100 ppb) con 5 niveles de concentración. Las muestras fueron cuantificadas utilizando la curva adecuada para cada caso. Como estándares internos se emplearon Ho y Rh en concentraciones de 10 ppb. El límite de detección (LOD) se calculó siguiendo el criterio de $3\sigma/m$ para 10 replicados del blanco, es decir tres veces la desviación estándar sobre la pendiente.

3.3.3. Vinos

Las muestras de vinos fueron analizadas mediante la técnica de ICP-MS empleando las siguientes condiciones instrumentales:

Tabla 3.5. Condiciones óptimas instrumentales de la determinación multielemental en las muestras de vinos.

Instrumento	Agilent 7900 ICP-MS (Tokyo, Japan) con SPS 4 autosampler
Flujo Ingreso muestra (mL min ⁻¹)	0,4
Introducción de la muestra	Nebulizador concéntrico. Cámara de nebulización de doble paso tipo Scott
Potencia RF (W)	1550
Flujo del gas (L min ⁻¹)	Plasma: 15; carrier: 1,09
Interfase	Conos de muestreo y skimmer de Ni
Adquisición de datos	Peak hopping
Dwell time (ms)	No se setea el dwelling time, si no que fijamos el número de barridos por réplica y el tiempo de integración. El número de barridos por réplica (sweeps per replicate) fue de 100
Flujo de gas (mL min ⁻¹) en la celda de colisión	4 [He]
Geometría de la celda	Octopolo
Octopolo RF (V)	200
Número de réplicas	3
Isotopos monitoreados	⁷ Li, ⁹ Be, ²⁷ Al, ⁵¹ V, ⁵² Cr, ⁵⁵ Mn, ⁵⁷ Fe, ⁵⁹ Co, ⁶⁰ Ni, ⁶³ Cu, ⁶⁶ Zn, ⁶⁹ Ga, ⁷⁵ As, ⁷⁸ Se, ⁸⁵ Rb, ⁸⁸ Sr, ⁹⁵ Mo, ¹⁰⁵ Pd, ¹⁰⁷ Ag, ¹¹¹ Cd, ¹¹⁸ Sn, ¹²¹ Sb, ¹³³ Cs, ¹³⁷ Ba, ²⁰¹ Hg, ²⁰⁵ Tl y ²⁰⁸ Pb.

El espectrómetro se calibró con las soluciones estándar externas certificadas. Para corregir las interferencias no espectrales, se utilizó como estándar interno para todas las determinaciones la solución "ICP-MS Internal std mix" (6-Li, Sc, Ge, Rh, In, Tb, Lu, Bi).

3.4. ANÁLISIS ESTADÍSTICO QUIMIOMÉTRICO

En la presente tesis doctoral, los resultados obtenidos a través del análisis multielemental en los distintos tipos de muestras se presentan mediante el cálculo de la media y desviación estándar de las concentración de los elementos en forma de tablas.

Para el tratamiento quimiométrico, los datos se dispusieron en matrices en las cuales las filas representaron las muestras y las columnas correspondieron a la concentración de los elementos determinados y en todos los casos una columna representó el origen geográfico de las muestras como variable categórica dependiente. Los concentraciones que se encontraron por debajo del LOD se trataron asumiendo los valores de LOD en los cálculos correspondientes para las técnicas quimiométricas. Para la etapa de preprocesamiento del conjunto de datos de

la matriz se empleó la estandarización debido a las amplias diferencias en la dimensionalidad de los datos obtenidos.

Los métodos estadísticos multivariantes permitieron verificar la contribución de cada variable al modelo y su capacidad para discriminar las diferentes categorías. La caracterización quimiométrica básica de las muestras de semillas, hollejos, jugos de uva y vinos investigadas se realizó mediante el análisis de componentes principales (PCA), que, como se expuso en el capítulo I, es una técnica no supervisada. Esta técnica se usó como un análisis exploratorio para dividir el conjunto de datos en un número menor de variables y proporcionar una representación más sencilla de toda la información. Otro método no supervisado utilizado en el caso de las muestras de los jugos de uva fue análisis de clúster o conglomerados (CA), dentro de este método se evaluaron dos algoritmos de agrupamiento, el análisis de agrupamiento jerárquico (HCA) y el agrupamiento de k-medias.

Finalmente, se utilizaron las diversas herramientas quimiométricas descritas en el capítulo I para evaluar los diferentes modelos para la clasificación de las muestras de acuerdo con su origen geográfico.

Todos los procedimientos estadísticos se llevaron a cabo utilizando el software estadístico R versión 3.3.0 y 3.6.0 con el paquete Caret para la clasificación de reconocimiento de patrones [1,2]. El paquete caret compila algoritmos de clasificación y regresión, contiene varias herramientas para desarrollar modelos predictivos utilizando los paquetes existentes en R y facilita el entrenamiento y optimización de los modelos. Principalmente es útil para la comparación de los algoritmos y también incluye métodos para el preprocesamiento de los datos, cálculo de la importancia de las variables y las visualizaciones de los modelos [3].

Además se empleó Infostat® software [4] para el análisis de varianza de prueba de Kruskal-Wallis atendiendo a los supuestos de que las variables mostraron una distribución sesgada (verificada por gráficos de probabilidad normales) y las varianzas no fueron homogéneas (verificadas por la prueba de Levene).

3.5. REFERENCIAS

- [1] R Core Team, R: A Language and Environment for Statistical Computing, (2016). <https://www.r-project.org/>.
- [2] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing., (2019). <https://www.r-project.org/>.
- [3] M. Kuhn, Building predictive models in R using the caret package, J. Stat. Softw. 28 (2008) 1–26.
- [4] R.C.W. Di Rienzo J.A., Casanoves F., Balzarini M.G., Gonzalez L., Tablada M., InfoStat versión 2016, (2016). <http://www.infostat.com.ar>.

CAPÍTULO IV: RESULTADOS Y DISCUSIÓN

4.1. CLASIFICACIÓN GEOGRÁFICA DE UVAS

4.1.1. Clasificación intra-regional de semillas de uva de Mendoza (Argentina) mediante herramientas quimiométricas y análisis multielemental

4.1.1.1. Introducción

Actualmente la identificación del origen geográfico de las uvas (*Vitis vinifera L.*) es un aspecto de gran interés para el comercio mundial, porque está directamente relacionado con la procedencia del vino y las prácticas de elaboración de una región específica. Del mismo modo, el conocimiento del origen de la uva es de gran importancia para los enólogos para asegurar la calidad del vino que se produce. En América del Sur, Argentina es uno de los principales productores y exportadores de vinos de alta calidad. Entre varias regiones de Argentina, la provincia de Mendoza es la principal región productora de uva y vino, ya que representa más del 75% de la producción nacional [1,2]. Las condiciones naturales de la provincia de Mendoza, como la altitud, los suelos áridos y rocosos y la amplitud térmica diferenciada entre el día y la noche, son ideales para el cultivo exitoso de diversas variedades de uvas de alta calidad (variedades blancas, rojas y rosadas) [3]. Sin embargo, además de la calidad, el hecho de asegurar el origen geográfico de las uvas es considerado un aspecto importante por los consumidores, y por lo tanto, de gran valor estratégico para las economías de las regiones que producen uvas y vinos. Dentro de este contexto, la trazabilidad de las uvas y vinos producidos en Argentina está siendo intensamente controlada para certificar origen y otorgar valor agregado a sus productos.

Últimamente se ha explorado ampliamente la clasificación geográfica de vinos de diferentes partes del mundo usando quimiometría. Entre los varios factores que están fuertemente correlacionados y afectan la composición del vino, puede ser mencionado la variedad o variedades de uvas, el lugar donde se cultivan las uvas, las condiciones meteorológicas, las prácticas agrícolas en los viñedos, y las técnicas utilizadas por los enólogos, entre otros. Sin embargo, la variable más influyente en el sabor del vino es la variedad de uva

debido a la composición química diferencial [4,5]. Las uvas presentan una matriz compleja compuesta por agua, azúcares, ácidos orgánicos, compuestos fenólicos, vitaminas, y minerales [6]. En general, la composición elemental del vino refleja la geoquímica del área de cultivo, la composición química de las materias primas, así como los factores de producción tales como las prácticas agrícolas, o procesos de vinificación [7]. Los elementos inorgánicos tienen varias ventajas como marcadores químicos, ya que no se metabolizan y permanecen inalterados o transformados durante los procesos enológicos [8,9]. Teniendo en cuenta este aspecto, el estudio del perfil mineral junto con el uso de herramientas quimiométricas podría utilizarse para la clasificación de las uvas. De hecho, podría ser una poderosa estrategia para discriminar vinos y otros productos de uva provenientes de diferentes orígenes geográficos. Además, los elementos traza en uvas se encuentran principalmente en las semillas [10,11], que también son muy importante porque son los principales subproductos derivados de la industria del vino, representando el 15% de los residuos sólidos [12,13].

Las técnicas de huella dactilar o fingerprint, combinan análisis químico con un análisis estadístico multivariado o métodos de minería de datos para resolver problemas complejos que requieren enfoques multivariados [14–16]. Para generar datos fiables sobre el contenido mineral de diferentes tipos de muestras, en general se usan técnicas de espectrometría de masas. Actualmente, ICP-MS es la técnica más utilizada debido a su alta sensibilidad y robustez para las determinaciones multi-elementales [17]. Sin embargo, aunque varios trabajos han reportado la aplicación de ICP-MS junto con quimiometría para discriminar vinos, la clasificación de las uvas según su origen geográfico no ha sido ampliamente estudiada [7,18,19]. Según el relevamiento de la literatura, actualmente no existen estudios en los que se hayan utilizado semillas de uva como huella dactilar de las uvas, ya que generalmente, todo el grano de uva se analiza sin discriminar qué parte podría ser más apropiada para la clasificación quimiométrica.

Por consiguiente, el propósito del actual trabajo fue evaluar el uso potencial de varias herramientas estadísticas multivariadas combinadas con un análisis multielemental realizado mediante la técnica de ICP-MS de muestras semillas de uva para la diferenciación de uvas según su origen geográfico. El potencial de esta estrategia fue probado para discriminar uvas cultivadas en regiones vitivinícolas cercanas dentro de la provincia de Mendoza, lo cual lo diferenció de otros trabajos que informan sobre la clasificación de uvas cultivadas en regiones de larga distancia, donde se podría esperar diferencias significativas en la composición del suelo y por lo tanto en la composición química de las uvas para lograr una fácil clasificación. Finalmente, este es el primer trabajo que informa la clasificación de uvas cultivadas en la provincia de Mendoza de acuerdo a su origen geográfico.

4.1.1.2. Determinación multielemental en semillas de uva

En la Tabla 4.1 se muestra la concentración de los diferentes elementos analizados mediante ICP-MS en las muestras de semilla de uva de distinta procedencia de origen, expresadas como valores medios con su correspondiente desviación estándar.

Tabla 4.1. Concentración multielemental de las muestras de semilla de uva según su origen geográfico (las concentraciones se expresan en $\mu\text{g/g}$ de base de materia seca).

Elemento	LOD ^a ($\mu\text{g/L}$)	Rivadavia (n = 92)		San Martín (n = 69)		Guaymallén (n = 25)		Junín (n = 29)		Maipú (n = 193)	
		Media	DE	Media	DE	Media	DE	Media	DE	Media	DE
Ag	1,11	0,074	0,042	0,040	0,018	0,25	0,31	0,21	0,19	0,067	0,087
As	0,48	0,076	0,091	0,10	0,05	0,70	0,96	0,25	0,28	0,19	0,37
Ce	0,45	0,15	0,01	<LOD	<LOD	0,016	0,011	0,041	0,013	0,029	0,037
Co	0,48	0,013	0,008	0,016	0,017	0,079	0,102	0,034	0,036	0,013	0,020
Cs	0,12	0,009	0,015	0,003	0,002	0,038	0,023	0,017	0,015	0,018	0,002
Cu	0,90	13,0	3,3	12,0	2,7	9,0	2,3	12,9	2,1	12,4	3,2
Eu	0,18	0,003	0,006	0,007	0,008	0,14	0,03	0,052	0,056	0,010	0,024
Fe	2,73	44,2	11,6	37,5	10,7	39,3	7,1	38,7	11,7	39,5	12,2
Ga	0,33	0,024	0,017	0,016	0,014	0,13	0,13	0,053	0,045	0,036	0,057
Gd	0,57	0,004	0,007	0,009	0,007	0,24	0,07	0,058	0,064	0,015	0,043
La	0,24	0,037	0,030	0,024	0,017	0,10	0,06	0,057	0,038	0,025	0,021
Lu	0,03	0,003	0,005	0,003	0,003	0,042	0,007	0,016	0,015	0,005	0,010
Mn	0,27	18,2	4,4	19,6	4,7	18,4	4,1	20,3	4,2	15,7	6,7
Mo	0,36	2,4	0,7	2,4	0,6	3,8	2,5	6,9	4,0	4,3	3,1
Nb	0,15	0,094	0,142	0,16	0,16	0,12	0,10	0,30	0,38	0,21	0,33
Nd	1,38	0,030	0,055	0,011	0,011	0,25	0,05	0,090	0,078	0,034	0,078
Ni	0,27	3,4	2,8	3,4	3,9	2,9	1,4	5,0	5,1	3,2	2,2
Pr	0,18	0,006	0,011	0,012	0,008	0,093	0,026	0,042	0,040	0,014	0,025
Rb	0,42	1,9	1,0	2,0	1,2	6,2	6,8	5,2	8,9	3,4	2,5
Sm	0,78	0,004	0,007	0,009	0,007	0,19	0,06	0,034	0,043	0,012	0,036
Te	2,85	1,2	1,3	2,0	1,9	22,5	13,9	3,6	5,5	2,0	4,0

Ti	3,45	3,0	1,3	2,6	1,2	3,3	1,9	3,1	1,0	2,6	1,1
Tl	0,03	0,093	0,047	0,12	0,08	0,18	0,05	0,15	0,03	0,11	0,07
Tm	0,06	0,007	0,009	0,004	0,003	0,040	0,012	0,015	0,012	0,007	0,011
U	0,06	0,047	0,020	0,010	0,006	0,021	0,007	0,036	0,012	0,023	0,023
V	0,42	0,017	0,010	0,008	0,004	0,15	0,06	0,032	0,028	0,033	0,046
Y	0,18	0,009	0,020	0,014	0,012	0,11	0,02	0,04	0,05	0,009	0,019
Zn	0,18	18,5	6,3	17,6	7,2	16,5	3,4	18,0	10,2	16,8	5,8
Zr	0,51	0,14	0,05	0,16	0,07	0,21	0,14	0,20	0,12	0,17	0,09

^aLOD: límite de detección instrumental

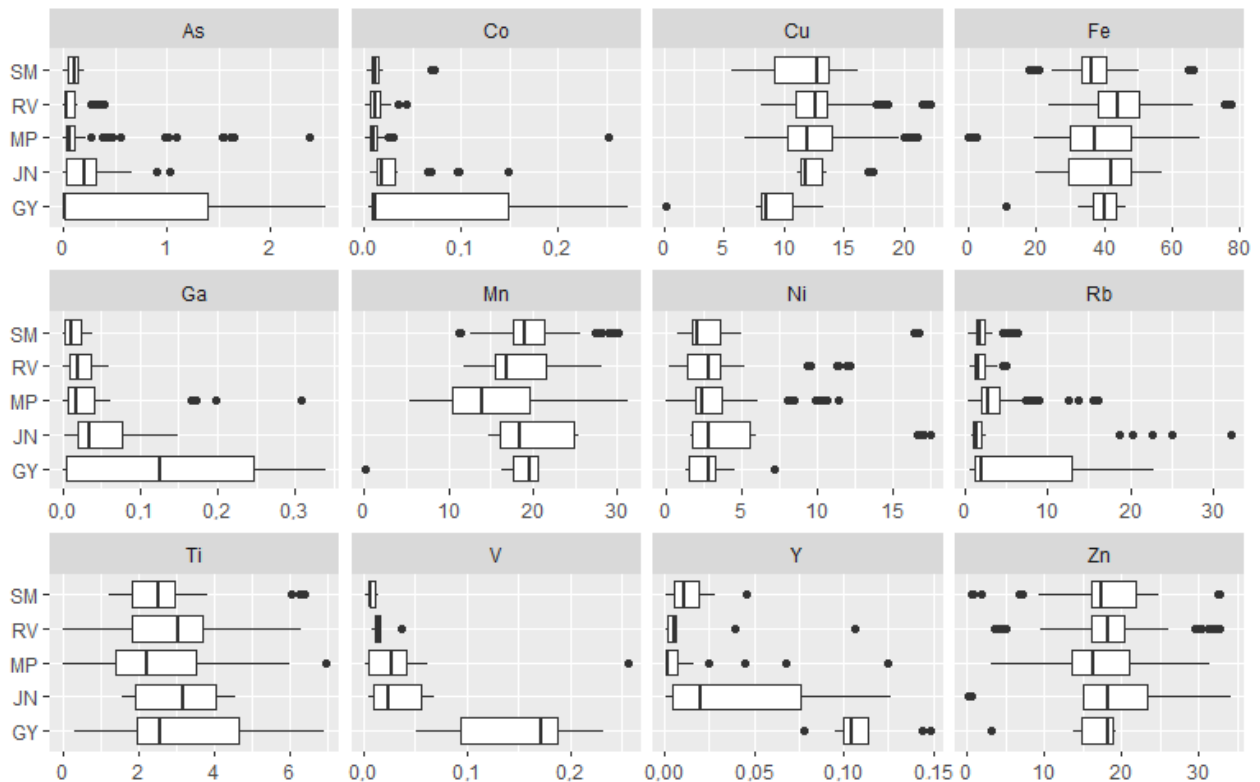
Del análisis de los resultados presentados en la Tabla 4.1, se observa que Fe es el elemento más abundante encontrado en las semillas de uva, seguido por Mn, Zn, y Cu. El contenido medio de Fe más bajo se obtuvo en San Martín (37,5 µg/g) mientras que la concentración más alta de este elemento se encontró en Rivadavia (44,2 µg/g). La concentración de Mn hallada estuvo en el rango de 15,7 µg/g y 20,3 µg/g, mientras que las de Zn y Cu estuvieron en el rango de 16,5 – 18,5 µg/g y 9 – 13 µg/g, respectivamente. Estas concentraciones estuvieron en concordancia con las reportadas por otros autores en las muestras de semillas, en general en el mismo orden de magnitud [12,20,21]. Con respecto a otros trabajos sobre semillas de uva, se encontraron diferencias en los niveles de estos elementos [11,22,23]. Estos minerales son considerados esenciales y determinan que las semillas de uva sean una excelente fuente de bioelementos, ya que desempeñan un papel importante en el normal funcionamiento del cuerpo humano. Por ejemplo, el Fe, entre otras funciones, está asociado con la producción de células sanguíneas, y Zn es esencial para el sistema inmunológico. Ambos nutrientes también se consideran potentes antioxidantes. Mientras tanto, el Mn es un elemento esencial tanto para los animales como para las plantas, y su deficiencia da como resultado anomalías esqueléticas y reproductivas graves en los mamíferos [13,24].

Los elementos encontrados a niveles de microgramos fueron Mo, Ni, Rb, Te y Ti, mientras que Ag, As, Co, Cs Ga, Nb, Tl, U, V y Zr se hallaron en concentraciones por debajo del rango del microgramo. Para estos elementos, se tiene que mencionar que no hay suficientes datos disponibles sobre su contenido en semillas de uva en la literatura para la comparación con los resultados obtenidos.

Por otro lado, los elementos llamados tierras raras (REE por su sigla en inglés) analizados, Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tm y Lu se encontraron en niveles ultra-traza o no se detectaron en algunas muestras de semilla de uva, esto está de acuerdo con el hecho de que

las semillas no son buenos reservorios de este tipo de elementos, como se puede observar en varios estudios, en los que sólo el 3% de las REE se acumulan en las semillas de uva [21,25].

La comparación de las concentraciones de los elementos analizados en las muestras de semillas de uva teniendo en cuenta el origen geográfico, muestra que las semillas presentan perfiles de concentración similar de los elementos según la región. Para poder tener una mejor visualización de los datos por región, se presenta en la Figura 4.1 los gráficos de caja y bigotes (box and whisker plot en inglés) que permiten comparar las distribuciones de las concentraciones de los diferentes elementos analizados en las semillas de las cinco regiones estudiadas. Se observa una tendencia de mayores concentraciones de los REE, como Eu, Gd, La, Lu, Nd, Pr, Sm, Tm e Y, como así también de los elementos minoritarios, Ag, As, Co, Cs, Ga, Tl, V y Zr en las muestras de semillas procedentes de la región de Guaymallén (GY).



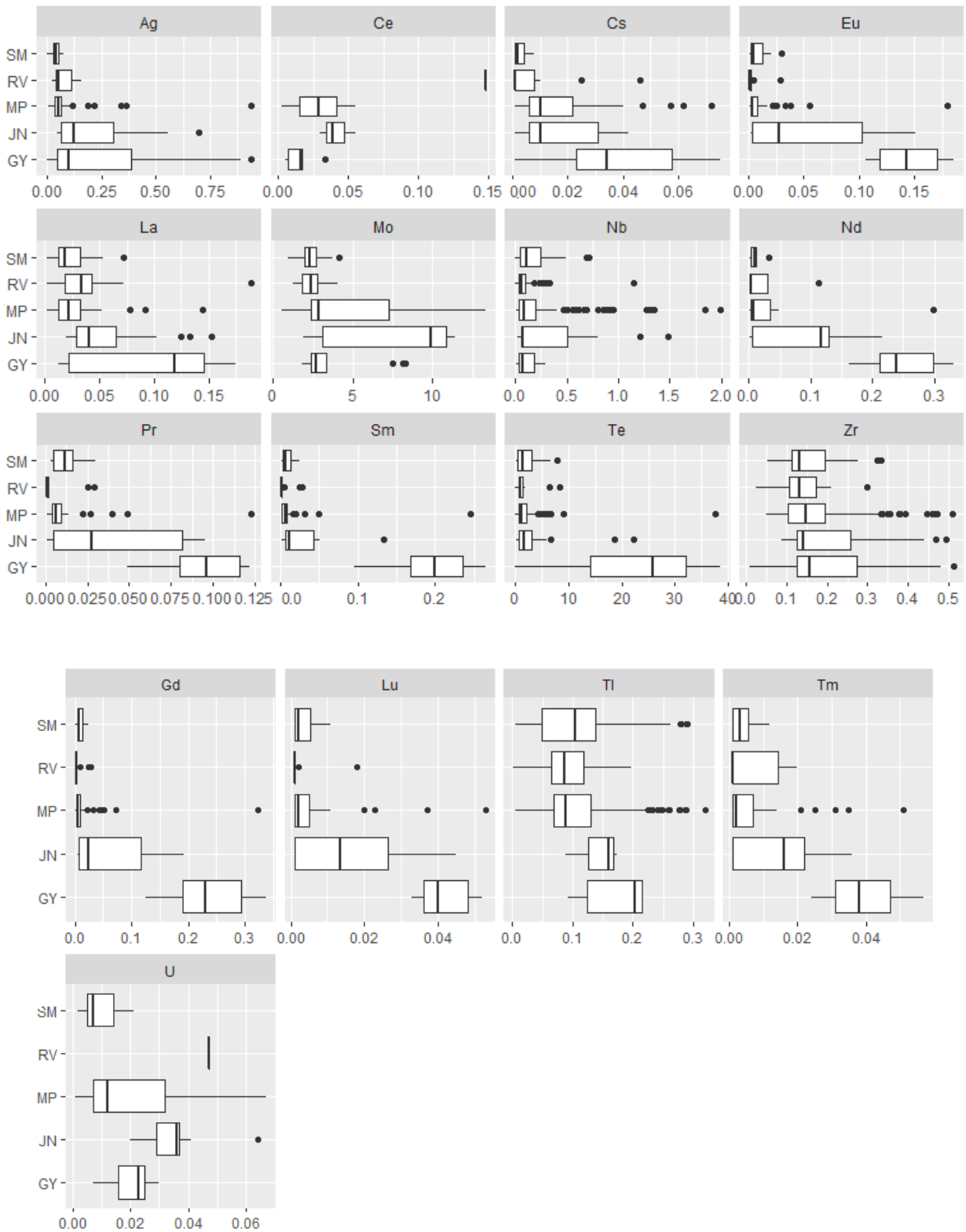


Figura 4.1. Gráficos Box-plot de los elementos determinados en semillas de uva según su región de procedencia: San Martín (SM), Rivadavia (RV), Maipú (MP), Junín (JN) y Guaymallén (GY).

4.1.1.3. Clasificación geográfica de semillas de uva de la provincia de Mendoza

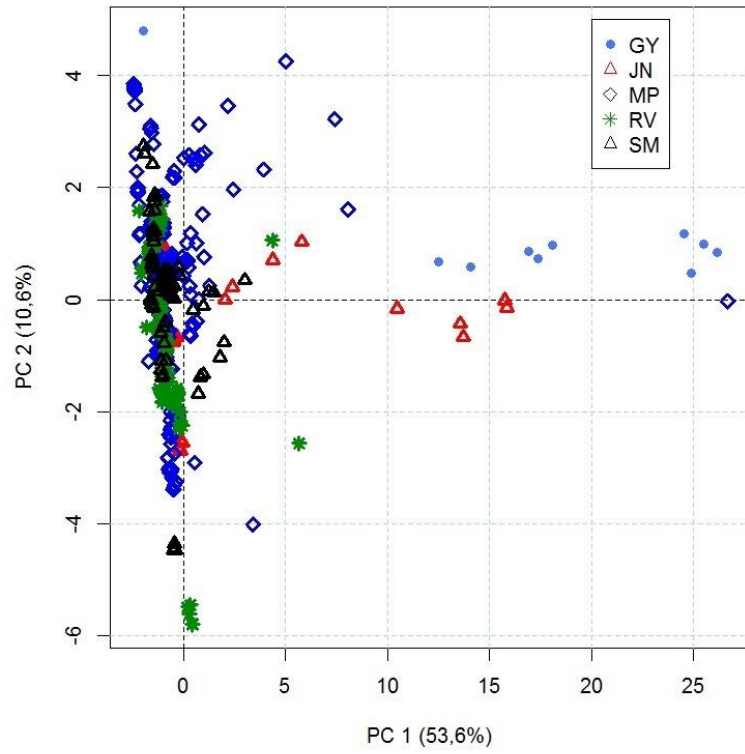
4.1.1.3.1. Análisis estadístico exploratorio

Para el análisis estadístico exploratorio, se realizó la caracterización quimiométrica básica de las semillas de uva investigadas mediante el método de PCA. Como se mencionó anteriormente, este método permite la determinación de las principales fuentes de variabilidad en el conjunto de datos, y muestra la agrupación natural de los objetos estudiados (muestras de semilla de uva) en el plano o en un espacio en 3D de las componentes principales más importantes, que son creadas por un combinación lineal no correlacionada de todas las variables originales. Por lo tanto, permite distinguir entre los objetos estudiados, según la región, y define las causas de la variabilidad [26].

Para llevar a cabo el análisis de PCA, se tuvieron en cuenta los datos de concentración de los 29 elementos analizados. El contenido total de información dado por los componentes principales se expresa en forma de porcentaje acumulado (cum. %) de la varianza total. Por ejemplo, cuando se utilizaron 29 variables, las primeras dos componentes principales (PC) representaron 64,2% de la variabilidad explicada. La primera componente principal (PC1) representó el 53,6% de la varianza total y la siguiente componente principal, 10,6% (PC2).

En la Figura 4.2 se muestran los dos gráficos más importantes de PCA, el de observaciones (scores) y variables (loadings) de las PC1 vs PC2, donde se muestra la agrupación natural de las 408 muestras de semilla de uva y la orientación de las variables respectivamente. La Fig. 4.2a muestra el gráfico de scores o puntuaciones de las dos primeras componentes principales. Se puede observar una gran superposición entre los scores correspondientes a las diferentes muestras identificadas según su origen. Sin embargo, las muestras de los departamentos Guaymallén (GY) y Junín (JN) mostraron puntuaciones positivas en el PC1 y pueden ser diferenciadas del resto de los grupos que tienen puntuaciones negativas en PC1.

a)



b)

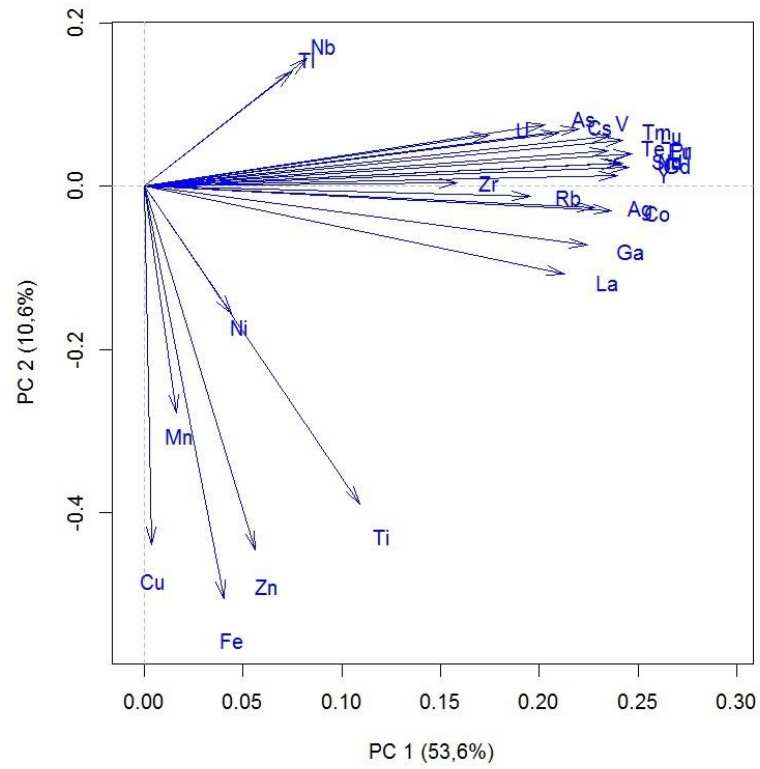


Figura 4.2. Gráfico PCA de a) observaciones y b) variables de la PC1 vs. PC2 de las semillas.

La Figura 4.2b presenta el gráfico de loadings en el plano PC1 vs.PC2, que ilustra la orientación de las variables (concentraciones de elementos) con respecto a las componentes principales más informativas PC1 y PC2. Se observa que la primera componente principal está fuertemente asociada con los valores de los elementos: Eu, Gd, Pr, Tm, Lu, Nd, Sm, Y y Co, indicando concentraciones más altas en las muestras que mostraron puntuaciones positivas (muestras GY). Por otro lado, Fe, Cu, Zn, Ti y Mn fueron las variables dominantes en la segunda componente principal, mientras que las muestras con puntuaciones negativas en PC2 corresponden a altas concentraciones de estos elementos y las observaciones positivas a altas concentraciones de Nb y Tl.

En resumen, los resultados obtenidos por PCA muestran que sólo algunas muestras de semillas de las regiones GY y JN pudieron diferenciarse considerando sus perfiles de elementos traza, pero otras muestras de semillas con diferentes orígenes geográficos no pudieron resolverse con este método quimiométrico no supervisado. Estos hallazgos sugirieron la necesidad de aplicar algoritmos quimiométricos más complejos para lograr la clasificación geográfica de las muestras.

4.1.1.3.2. Análisis estadístico clasificatorio

Las técnicas de reconocimiento de patrones supervisadas, implican principalmente la clasificación de objetos en grupos o categorías creando reglas de clasificación [27]. Se aplicó un método supervisado clásico como el LDA considerando todas las muestras de semillas de uva y se analizó su exactitud o accuracy. Las dos primeras funciones canónicas discriminantes (DFs) explicaron un 78,2 % de la varianza. La Figura 4.3 muestra los patrones de distribución de las muestras de semillas de uva según su origen geográfico en el gráfico definido por las dos funciones discriminantes. Esta figura muestra un notable solapamiento entre los grupos, distinguiéndose la presencia de 3 grupos principales: Guaymallén (GY) (scores negativos en DF1 y DF2), Junín (JN) (scores positivos en DF1 y negativos en DF2) y el tercer grupo formado por Maipú (MP) + Rivadavia (RV) + San Martín (SM) en el centro de la gráfica.

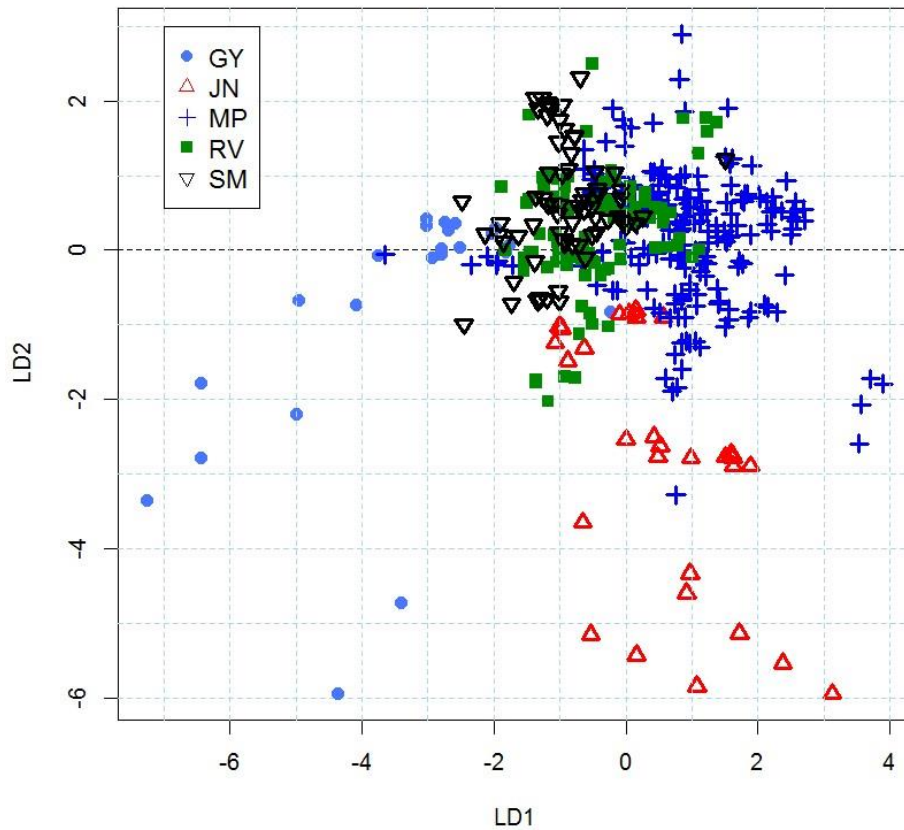


Figura 4.3. Gráfico de dispersión de las dos primeras funciones discriminantes del LDA de muestras de semilla de uva según su origen geográfico.

Las accuracies o exactitud medias de predicción para las 408 muestras fue de 91,6% para semillas de GY; 95% para muestras de JN; 78,3% para muestras de MP; 60,5% para muestras de RV y 55,1% para muestras de SM. La mayoría de las muestras de las regiones de GY y JN se clasificaron correctamente como se predijo a partir de la figura anterior. Las principales dificultades surgieron con las muestras de las localidades restantes (MP, RV y SM). Métodos no lineales fueron necesarios para resolver el problema de la predicción del origen geográfico de las muestras de semillas de uva.

Con la finalidad de realizar el análisis de clasificación predictivo, la matriz de datos de las concentraciones de los elementos, se dividió aleatoriamente en un conjunto de entrenamiento (70% de los objetos de toda la matriz de datos), con muestras de clase

conocidas para calcular un clasificador, y el conjunto de prueba o testeo (30%) que contiene las muestras restantes no incluidas en el entrenamiento y que sirven para validar el rendimiento de los modelos.

En el muestreo aleatorio, la división en subgrupos se realizó de manera estratificada, es decir, se mantuvo la proporción de cada clase en la matriz original en los nuevos subgrupos. Dado que los resultados obtenidos estuvieron condicionados por el punto en el que se realizó la división de la matriz de datos, esta operación se repitió n veces para obtener valores promedio para la tasa de éxito, permitiendo la comparación entre los diferentes métodos discriminatorios analizados. Los casos incluidos en cada conjunto se cambiaron aleatoriamente para cada modelo reproducido.

Se seleccionaron y probaron cinco modelos quimiométricos, LDA, PLS-DA, k-NN, SVM y RF, para clasificar las muestras de semillas de uva de acuerdo con su origen geográfico. Los métodos PLS-DA, k-NN, SVM y RF, requirieron optimizar algunos parámetros para construir el modelo antes de su evaluación como herramienta de predicción geográfica. Cada uno de los modelos fue entrenado usando una validación cruzada de k iteraciones o k -fold cross-validation (repetida n veces) en el conjunto de entrenamiento para construir los diferentes clasificadores. Este procedimiento se repitió n veces, por lo que cada subconjunto se utilizó para probar al menos una vez. En este trabajo se utilizó $k = 10$ y $n = 5$.

Los parámetros optimizados para cada modelo fueron: el número de componentes significativas ($ncomp$) para PLS-DA; el número de vecino k para k-NN; el factor de penalización C , ϵ de ϵ -insensitive loss función y el tipo de kernel para SVM, y finalmente el número de variables evaluadas en cada división (muestreo estratificado) ($mtry$) y el número de árboles (nt) para RF, todos calculados mediante el uso de la técnica de validación cruzada descrita anteriormente teniendo en cuenta la tasa máxima de exactitud como criterio para su selección. Una vez que se seleccionaron los valores óptimos para cada modelo, la sensibilidad (muestras pertenecientes a una clase y clasificadas correctamente en esta clase), especificidad (muestras

que no pertenecen a una clase y clasificadas correctamente como no pertenecientes), y la tasa de exactitud general fueron los parámetros propuestos para la evaluación de los modelos de clasificación supervisados en el conjunto de datos de prueba o testeo. En la Tabla 4.2. se presentan los resultados que indican el rendimiento de los diferentes métodos de clasificación.

A partir del análisis de la tabla 4.2, se observa que los cinco métodos quimiométricos evaluados mostraron diferentes grados de éxito en la predicción de las muestras de testeo o prueba. El orden de las tasas de predicción fue el siguiente: RF> SVM > k-NN> LDA> PLS-DA. Los modelos lineales, como LDA y PLS-DA, presentaron el peor desempeño desde el punto de vista de accuracy general, por lo que el uso de métodos no lineales resolvió el problema de clasificación, siendo RF el modelo ideal para discriminar las muestras de semilla de uva según su origen geográfico, con una precisión de clasificación global del 98,3%. Por lo tanto, en base a los resultados, se obtuvo una buena discriminación de las muestras en estudio con modelos no lineales, que puede explicarse por la flexibilidad y la capacidad del algoritmo para crear un modelo generalizado, incluso para grupos de entrenamiento pequeños [28].

Finalmente, se realizó una comparación adicional del desempeño de los cinco métodos de clasificación aplicados mediante el uso de la repetición del estimador de un tercio de reserva. (one-third holdout). Dado que el resultado depende de la elección correcta de los datos de prueba, la matriz de datos fue dividida 30 veces y los resultados se compararon entre sí. Este procedimiento proporcionó una distribución de la exactitud general para comparar el rendimiento de cada método en su configuración de parámetros óptima. En la Figura 4.4, se observa las distribuciones representadas por gráficos de caja y bigotes. Los límites alrededor de la media representan la amplitud de la dispersión de los resultados obtenidos para cada ciclo de remuestreo. Los resultados obtenidos por este enfoque confirman que el mejor método para la clasificación geográfica de este conjunto de datos fue RF, con una precisión media de casi el 97% y un valor kappa de 11.

Tabla 4.2. Resultados de la clasificación obtenidos con los diferentes modelos evaluados.

Grupos	Número de muestras		LDA		PLS-DA (<i>ncomp</i> = 3) ^a		k-NN (<i>k</i> = 5) ^b		SVM (<i>C</i> = 32; ϵ = 0,09) ^c		RF (<i>nt</i> = 500; <i>mtry</i> = 11) ^d	
	Entrena- -miento	Testeo	Sens. (%)	Espec. (%)	Sens. (%)	Espec. (%)	Sens. (%)	Espec. (%)	Sens. (%)	Espec. (%)	Sens. (%)	Espec. (%)
GY	18	7	71	100	57	100	71	100	43	100	100	100
JN	20	9	62	98	0	100	87	95	100	100	100	100
MP	135	58	80	77	89	30	91	92	98	89	98	98
RV	64	28	40	77	26	90	85	95	89	100	96	100
SM	48	21	40	92	0	95	75	97	100	99	100	99
accuracy media (%)			63,0		52,1		85,7		93,3		98,3	

^a *ncomp*: número de componentes significativas.

^b *k*: número de *k* vecinos.

^c *C*: factor de penalidad; ϵ : ϵ -insensitive loss function.

^d *nt*: número de árboles; *mtry*: número de variables probadas en cada división.

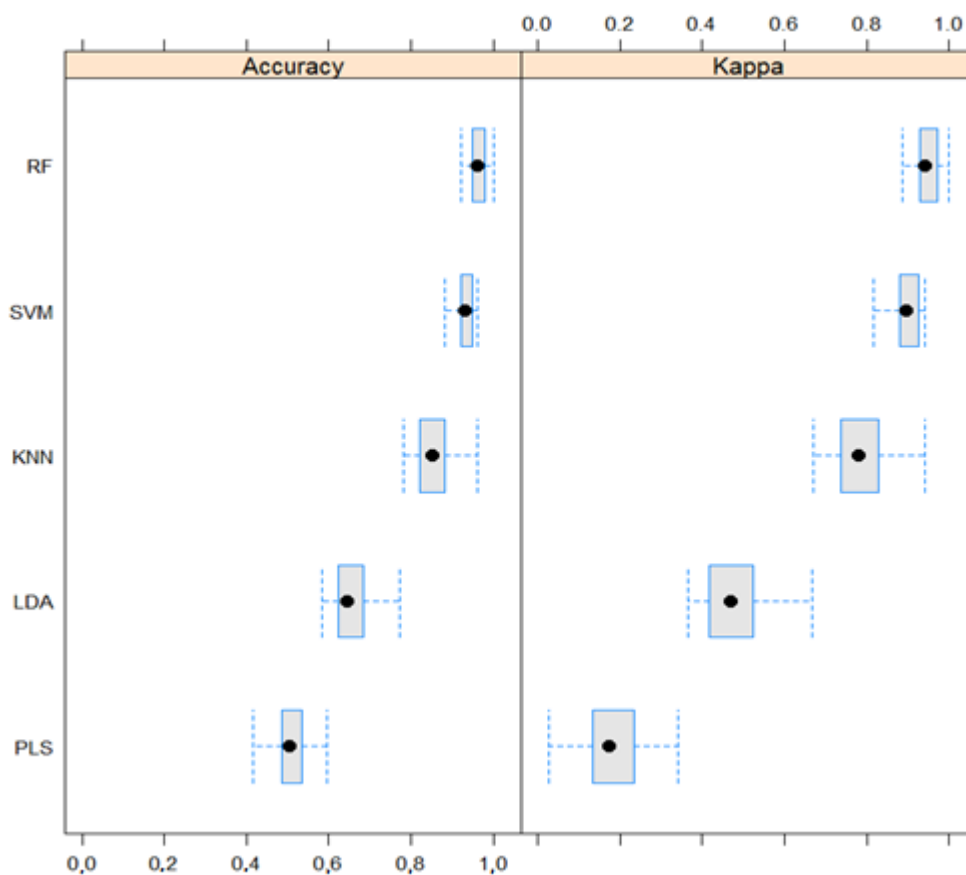


Figura 4.4. Gráfico de caja y bigotes comparando los modelos aplicados para la clasificación de semillas de uva.

4.1.1.4. Conclusiones

La primera clasificación intrarregional de uvas producidas en diferentes viñedos de la provincia de Mendoza (Argentina) basada en el análisis multielemental de semillas de uva se logró en el presente trabajo. La aplicación de la técnica de análisis multielemental por ICP-MS, combinada con herramientas quimiométricas, tales como RF, proporcionó una herramienta robusta para la clasificación geográfica de semillas de uvas producidas en los diferentes lugares de esta provincia vitivinícola. Los resultados de la predicción tanto en las muestras de entrenamientos como en las de testeo, demostraron que el modelo RF propuesto es un método eficaz y eficiente para la clasificación de muestras de semillas de uva. Además, 500 árboles y 29

variables fueron optimizados y seleccionados como los mejores parámetros para la clasificación de la muestra de semillas de uvas. Los modelos predictivos para las muestras de semillas de uva contienen información sobre su composición química que contribuye a la clasificación de diferentes orígenes geográficos. Los modelos evaluados LDA, PLS-DA, k-NN, SVM y RF para la predicción del origen geográfico, mostraron que los métodos de SVM y RF tuvieron una alta tasa de accuracy de predicción. De hecho, la tasa promedio de exactitud de la predicción del método de RF fue más alta que los otros métodos. Por lo tanto, se demuestra que el método de RF es una herramienta prometedora para la clasificación y el control de calidad de las materias primas utilizadas en la industria vitivinícola.

4.1.1.5. Referencias

- [1] Instituto Nacional de Vitivinicultura (INV), Informe anual de cosecha y elaboración 2018, (2018). www.inv.gob.ar (accessed October 10, 2019).
- [2] V. Castex, E.M. Tejada, M. Beniston, Water availability, use and governance in the wine producing region of Mendoza, Argentina, *Environ. Sci. Policy*. 48 (2015) 1–8. doi:10.1016/j.envsci.2014.12.008.
- [3] Fundación ProMendoza, (2016). <http://www.promendoza.com/> (accessed June 15, 2016).
- [4] A.G. Reynolds, 11 – Viticultural and vineyard management practices and their effects on grape and wine quality, in: *Manag. Wine Qual.*, 2010: pp. 365–444. doi:10.1533/9781845699284.3.365.
- [5] C. van Leeuwen, 9 – Terroir: the effect of the physical environment on vine growth, grape ripening and wine sensory attributes, in: *Manag. Wine Qual.*, 2010: pp. 273–315. doi:10.1533/9781845699284.3.273.
- [6] R.S. Jackson, 6 - Chemical Constituents of Grapes and Wine BT - Wine Science (Fourth Edition), in: *Food Sci. Technol.*, Academic Press, San Diego, 2014: pp. 347–426.

doi:<http://dx.doi.org/10.1016/B978-0-12-381468-5.00006-3>.

- [7] A. Versari, V.F. Laurie, A. Ricci, L. Laghi, G.P. Parpinello, Progress in authentication, typification and traceability of grapes and wines by chemometric approaches, *Food Res. Int.* 60 (2014) 2–18. doi:[10.1016/j.foodres.2014.02.007](https://doi.org/10.1016/j.foodres.2014.02.007).
- [8] P. Pohl, What do metals tell us about wine?, *TrAC - Trends Anal. Chem.* 26 (2007) 941–949. doi:[10.1016/j.trac.2007.07.005](https://doi.org/10.1016/j.trac.2007.07.005).
- [9] J. Saurina, Characterization of wines using compositional profiles and chemometrics, *TrAC - Trends Anal. Chem.* 29 (2010) 234–245. doi:[10.1016/j.trac.2009.11.008](https://doi.org/10.1016/j.trac.2009.11.008).
- [10] M.P. Fabani, M.E. Toro, F. Vázquez, M.P. Díaz, D.A. Wunderlin, Differential absorption of metals from soil to diverse vine varieties from the valley of tulum (Argentina): Consequences to evaluate wine provenance, *J. Agric. Food Chem.* 57 (2009) 7409–7416. doi:[10.1021/jf901572k](https://doi.org/10.1021/jf901572k).
- [11] S.Y. Rogiers, D.H. Greer, J.M. Hatfield, B.A. Orchard, M. Keller, Mineral sinks within ripening grape berries (*Vitis vinifera* L.), *Vitis - J. Grapevine Res.* 45 (2006) 115–123.
- [12] J. Lachman, A. Hejtmánková, K. Hejtmánková, Š. Horníčková, V. Pivec, O. Skala, M. Dědina, J. Příbyl, Towards complex utilisation of winemaking residues: Characterisation of grape seeds by total phenols, tocopherols and essential elements content as a by-product of winemaking, *Ind. Crops Prod.* 49 (2013) 445–453. doi:[10.1016/j.indcrop.2013.05.022](https://doi.org/10.1016/j.indcrop.2013.05.022).
- [13] S. Mironeasa, A. Leahu, G. Codina, Grape Seed: physico-chemical, structural characteristics and oil content, *J. Agroalim. Process. Technol.* 16 (2010) 1–6.
- [14] F. Galgano, F. Favati, M. Caruso, T. Scarpa, A. Palma, Analysis of trace elements in southern Italian wines and their classification according to provenance, *LWT - Food Sci. Technol.* 41 (2008) 1808–1815. doi:[10.1016/j.lwt.2008.01.015](https://doi.org/10.1016/j.lwt.2008.01.015).
- [15] M.R. González-Centeno, C. Rosselló, S. Simal, M.C. Garau, F. López, A. Femenia, Physico-

- chemical properties of cell wall materials obtained from ten grape varieties and their byproducts: Grape pomaces and stems, *LWT - Food Sci. Technol.* 43 (2010) 1580–1586. doi:10.1016/j.lwt.2010.06.024.
- [16] L. Mutihac, R. Mutihac, Mining in chemometrics, *Anal. Chim. Acta.* 612 (2008) 1–18. doi:10.1016/j.aca.2008.02.025.
- [17] M. Tanner, D. Günther, Short transient signals, a challenge for inductively coupled plasma mass spectrometry, a review, *Anal. Chim. Acta.* 633 (2009) 19–28. doi:10.1016/j.aca.2008.11.041.
- [18] A. Cugnetto, L. Santagostini, L. Rolle, S. Guidoni, V. Gerbi, V. Novello, Tracing the “terroirs” via the elemental composition of leaves, grapes and derived wines in cv Nebbiolo (*Vitis vinifera* L.), *Sci. Hortic. (Amsterdam)*. 172 (2014) 101–108. doi:10.1016/j.scienta.2014.03.055.
- [19] R. Ferrer-Gallego, J.M. Hernández-Hierro, J.C. Rivas-Gonzalo, M.T. Escribano-Bailón, A comparative study to distinguish the vineyard of origin by NIRS using entire grapes, skins and seeds, *J. Sci. Food Agric.* 93 (2013) 967–972. doi:10.1002/jsfa.5842.
- [20] S.G. Tangolar, Y. Ozoğul, S. Tangolar, A. Torun, Evaluation of fatty acid profiles and mineral content of grape seed oil of some grape genotypes., *Int. J. Food Sci. Nutr.* 60 (2009) 32–39. doi:10.1080/09637480701581551.
- [21] Y. Yang, C. Duan, H. Du, J. Tian, Q. Pan, Trace element and rare earth element profiles in berry tissues of three grape cultivars, *Am. J. Enol. Vitic.* 61 (2010) 401–407.
- [22] M.M. Ozcan, Mineral Contents of Several Grape Seeds, *Asian J. Chem.* 22 (2010) 6480–6488.
- [23] M. Spanghero, A.Z.M. Salem, P.H. Robinson, Chemical composition, including secondary metabolites, and rumen fermentability of seeds and pulp of Californian (USA) and Italian grape pomaces, *Anim. Feed Sci. Technol.* 152 (2009) 243–255.

doi:10.1016/j.anifeedsci.2009.04.015.

- [24] S.S. Mitić, M. V. Obradović, M.N. Mitić, D. a. Kostić, A.N. Pavlović, S.B. Tošić, M.D. Stojković, Elemental Composition of Various Sour Cherry and Table Grape Cultivars Using Inductively Coupled Plasma Atomic Emission Spectrometry Method (ICP-OES), *Food Anal. Methods*. 5 (2011) 279–286. doi:10.1007/s12161-011-9232-2.
- [25] D. Bertoldi, R. Larcher, G. Nicolini, M. Bertamini, G. Concheri, Distribution of rare earth elements in *Vitis vinifera* L. “Chardonnay” berries, *Vitis*. 48 (2009) 49–51.
- [26] D. Kruzlicova, Ž. Fiket, G. Kniewald, Classification of Croatian wine varieties using multivariate analysis of data obtained by high resolution ICP-MS analysis, *Food Res. Int.* 54 (2013) 621–626. doi:10.1016/j.foodres.2013.07.053.
- [27] R.G. Brereton, Pattern recognition in chemometrics, *Chemom. Intell. Lab. Syst.* 149 (2015) 90–96. doi:10.1016/j.chemolab.2015.06.012.
- [28] J.E. Gaiad, M.J. Hidalgo, R.N. Villafañe, E.J. Marchevsky, R.G. Pellerano, Tracing the geographical origin of Argentinean lemon juices based on trace element profiles using advanced chemometric techniques, *Microchem. J.* 129 (2016) 243–248. doi:10.1016/j.microc.2016.07.002.
- [29] International Organization of Vine and Wine (OIV), State of the sector in 2018, 2019. <http://www.oiv.int/public/medias/6679/en-oiv-state-of-the-vitiviniculture-world-market-2019.pdf>.

4.1.2. Técnicas de minería de datos basado en la composición química de hollejos de uva para la evaluación de la calidad y la predicción de la trazabilidad de las uvas

4.1.2.1. Introducción

La uva (*Vitis vinifera L.*) es una de las frutas más cultivadas en el mundo. En el año 2018 se cultivaron una superficie mundial de aproximadamente 7,4 millones de hectáreas, de las cuales, alrededor del 50% de la cosecha de uva se destinó a la producción de vino [1]. Como ya se expuso anteriormente, Argentina ocupa una posición relevante en el contexto vitivinícola internacional, siendo uno de los principales productores y exportadores de vino. En Argentina, la provincia de Mendoza es la principal región productora de uva y vino del país, representando más del 75% de su producción nacional [2]. Actualmente, la creciente demanda de alimentos de alta calidad con un origen geográfico certificado ha provocado que la industria del vino genere los sistemas de certificación de origen, que permiten garantizar la autenticidad, la calidad y añadir valor comercial a los diferentes vinos comercializados en todo el mundo.

La composición mineral refleja la geoquímica en la que se cultivan las vides y, como consecuencia, los productos de la vid están influenciados por la composición del suelo [3]. Las técnicas de huellas dactilares son una forma de alcanzar la clasificación geográfica, combinando el análisis químico con las herramientas de minería de datos. Existen varias estrategias analíticas para el análisis de múltiples elementos. En los últimos años, la espectrometría de masas de plasma acoplada inductivamente (ICP-MS) ha ganado popularidad como instrumento para realizar análisis de trazas de metales. Las razones principales pueden deberse a su precisión y rapidez para determinar elementos traza y ultra traza, además de su capacidad para manejar matrices simples y complejas con interferencias matriciales mínimas [4].

La minería de datos se define como el procedimiento de encontrar y describir patrones estructurales mediante el análisis de cantidades sustanciales de datos. Los patrones que se descubren durante la minería de datos permiten hacer predicciones sobre nuevos datos que se pueden utilizar para clasificar ejemplos desconocidos [5–7]. La minería de datos implica una amplia variedad de técnicas, que implementan el aprendizaje supervisado; lo que significa que se emplea un conjunto de datos de entrenamiento etiquetados para calcular una función capaz de mapear un conjunto de variables en una etiqueta de clase, y lograr una clasificación de un gran conjunto de objetos en las clases predefinidas. El proceso de "entrenamiento" de un modelo también es sinónimo de "aprendizaje", donde el "aprendizaje automático" es el procedimiento para descubrir las relaciones entre las variables predictoras y de respuesta utilizando enfoques estadísticos basados en computadora [8].

Los modelos de clasificación que se incluyen en la minería de datos presentan algoritmos que han sido ampliamente estudiados en el campo del aprendizaje automático [8]. Los algoritmos más comunes utilizados con éxito para los problemas de clasificación y predicción son los algoritmos basados en la lógica, como los árboles de decisión y los clasificadores basados en reglas; otros enfoques incluyen algoritmos estadísticos puros, como por ejemplo las redes bayesianas. Los modelos no lineales basados en redes neuronales y regresión no lineal también se usan ampliamente. Finalmente, las Máquinas de vectores de soporte (SVM por su sigla en inglés) y Bosques aleatorios (RF por su sigla en inglés) se reconocen también como algoritmos exitosos para resolver problemas de clasificación [9].

En el presente trabajo de tesis doctoral se presenta un estudio comparativo de algoritmos de minería de datos para la clasificación geográfica de muestras de hollejos de uva, basado en su composición mineral analizada por ICP-MS. Se evaluaron técnicas populares de minería de datos como máquinas de vectores de soporte (SVM), regresión logística multinomial (MLR), k vecinos

más cercanos (k-NN) y bosque aleatorio (RF), con el cual se evaluó importancia de las variables. Varios autores han centrado sus estudios de clasificación geográfica en vinos y, en menor medida, en muestras de uvas enteras, por lo que el propósito de este trabajo se enfocó en evaluar un subproducto de la vinificación para identificar el origen geográfico de las uvas de la región vitivinícola más importante de Argentina mediante el desarrollo de modelos de minería de datos.

4.1.2.2. Concentración multielemental en hollejos de uva

La Tabla 4.3 muestra los valores medios y de desviación estándar expresados en $\mu\text{g/g}$, de los elementos analizados en las muestras de hollejo de uva en función de su procedencia (Rivadavia, San Martín, Guaymallén, Junín y Maipú).

Los resultados del análisis de hollejos de uva mediante ICP-MS mostraron que Fe es el elemento más abundante en todas las muestras, seguido de Ni, Zn y Cu, los cuales presentaron concentraciones superiores a $10 \mu\text{g/g}$. Los niveles de concentración de estos elementos en las pieles de uva estuvieron de acuerdo con los reportados por otros autores [10–16]. Luego se encontraron, los elementos Mo, Te, Mn, Rb y Ti con concentraciones mayores al microgramo, mientras que Ag, As, Co, Cs, Ga, Nb, Tl, U, V y Zr se cuantificaron con concentraciones por debajo del rango de los microgramos. Finalmente, los elementos de tierras raras (REE), incluidos Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tm y Lu, se encontraron en niveles ultra-traza en todas las muestras. Dentro de estos REE, Ce, La y Nd mostraron las concentraciones más altas. Estos valores estuvieron de acuerdo con un trabajo previamente realizado por Bertoldi et. al., quienes informaron que las pieles de uva son los principales reservorios de estos REEs [17]. En resumen, los resultados de la composición de las muestras de hollejo de uva indicaron que el contenido mineral presenta perfiles de concentración similares según la procedencia geográfica diferente, sin embargo, se observó una ligera tendencia a mayores concentraciones de REEs, como Ce, La y Nd, en muestras de la región de Guaymallén.

Tabla 4.3. Concentración de los elementos determinados en las muestras de hollejo de uva según su origen geográfico (las concentraciones se expresan en $\mu\text{g/g}$ de base de materia seca).

Elemento	LOD ^a ($\mu\text{g/L}$)	Rivadavia (n = 90)		San Martín (n = 74)		Guaymallén (n = 22)		Junín (n = 27)		Maipú (n = 200)	
		Media	DE	Media	DE	Media	DE	Media	DE	Media	DE
Ag	1,11	0,094	0,040	0,14	0,05	0,14	0,04	0,15	0,04	0,13	0,06
As	0,48	0,078	0,060	0,17	0,17	0,13	0,14	0,10	0,08	0,10	0,09
Ce	0,45	0,039	0,016	0,039	0,018	0,22	0,14	0,045	0,018	0,042	0,029
Co	0,48	0,027	0,013	0,032	0,018	0,031	0,010	0,036	0,029	0,022	0,013
Cs	0,12	0,002	0,002	0,002	0,001	0,004	0,002	0,002	0,001	0,006	0,005
Cu	0,90	12,9	3,2	13,9	4,1	8,9	2,7	11,5	4,8	12,9	3,5
Eu	0,18	0,002	0,001	0,001	0,001	0,001	0,0003	0,001	0,0004	0,001	0,0006
Fe	2,73	27,4	9,2	25,5	9,8	29,4	8,2	30,4	19,1	25,2	10,3
Ga	0,33	0,017	0,006	0,018	0,008	0,021	0,008	0,019	0,007	0,016	0,008
Gd	0,57	0,002	0,002	0,002	0,001	0,003	0,002	0,002	0,001	0,002	0,001
La	0,24	0,036	0,020	0,036	0,033	0,19	0,12	0,039	0,015	0,047	0,057
Lu	0,03	0,002	0,002	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
Mn	0,27	2,5	0,7	3,3	1,1	2,6	0,8	3,4	1,3	2,5	1,0
Mo	0,36	4,6	3,1	4,1	1,9	6,1	2,5	5,5	2,9	3,9	2,3
Nb	0,15	0,13	0,13	0,23	0,37	0,21	0,36	0,18	0,21	0,14	0,19
Nd	1,38	0,008	0,005	0,008	0,004	0,011	0,006	0,010	0,005	0,009	0,004
Ni	0,27	13,3	9,3	12,9	6,7	17,8	9,5	16,5	18,9	9,6	7,5
Pr	0,18	0,003	0,002	0,002	0,001	0,005	0,002	0,003	0,002	0,002	0,001
Rb	0,42	2,4	1,3	2,2	0,6	1,9	0,3	1,5	0,5	3,6	1,4
Sm	0,78	0,002	0,001	0,002	0,001	0,003	0,001	0,003	0,001	0,002	0,001
Te	2,85	4,1	4,3	3,7	3,9	3,7	3,8	7,3	8,3	4,7	4,3
Ti	3,45	0,84	0,64	0,84	0,58	1,1	0,7	0,98	0,68	0,77	0,60

Tl	0,03	0,14	0,10	0,093	0,060	0,11	0,08	0,13	0,07	0,068	0,049
Tm	0,06	0,002	0,002	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
U	0,06	0,019	0,007	0,021	0,010	0,026	0,012	0,023	0,004	0,022	0,009
V	0,42	0,024	0,025	0,017	0,009	0,019	0,010	0,048	0,035	0,023	0,018
Y	0,18	0,006	0,002	0,006	0,001	0,008	0,003	0,008	0,002	0,006	0,002
Zn	0,18	13,4	9,7	14,9	10,9	14,6	5,6	9,0	4,0	10,1	4,3
Zr	0,51	0,39	0,36	0,45	0,33	0,53	0,39	0,46	0,37	0,46	0,36

^aLOD: límite de detección instrumental

Las uvas son una fuente importante de minerales con funciones fisiológicas relevantes para el organismo humano. Por ejemplo, en concentraciones naturales, el Fe es un elemento esencial y desempeña un papel crucial asociado a la producción de células sanguíneas entre otras funciones, el Zn es indispensable para el sistema inmune y el Cu es parte de muchas enzimas involucradas en procesos redox [15,18]. Sin embargo, a concentraciones elevadas, los minerales como Fe y Cu pueden ser indeseables para el proceso enológico debido a que causan alteraciones del sabor y turbidez (casse) en los vinos [19]. El consumo diario de esta fruta contribuye significativamente en los requerimientos del organismo humano para elementos esenciales. Por otro lado, se sabe que varios metales, como Pb, Cd y As, son potencialmente tóxicos. Los elementos analizados en el presente trabajo se encontraron en niveles naturales y de acuerdo con los reportados en otros trabajos publicados en la literatura. En cuanto a la concentración de As, los valores obtenidos en nuestro trabajo fueron inferiores al nivel residual máximo establecido en frutas por el Código Alimentario Argentino [20] (Secretaría de Estado de Salud Pública, 2017) y las Directivas de la UE [21].

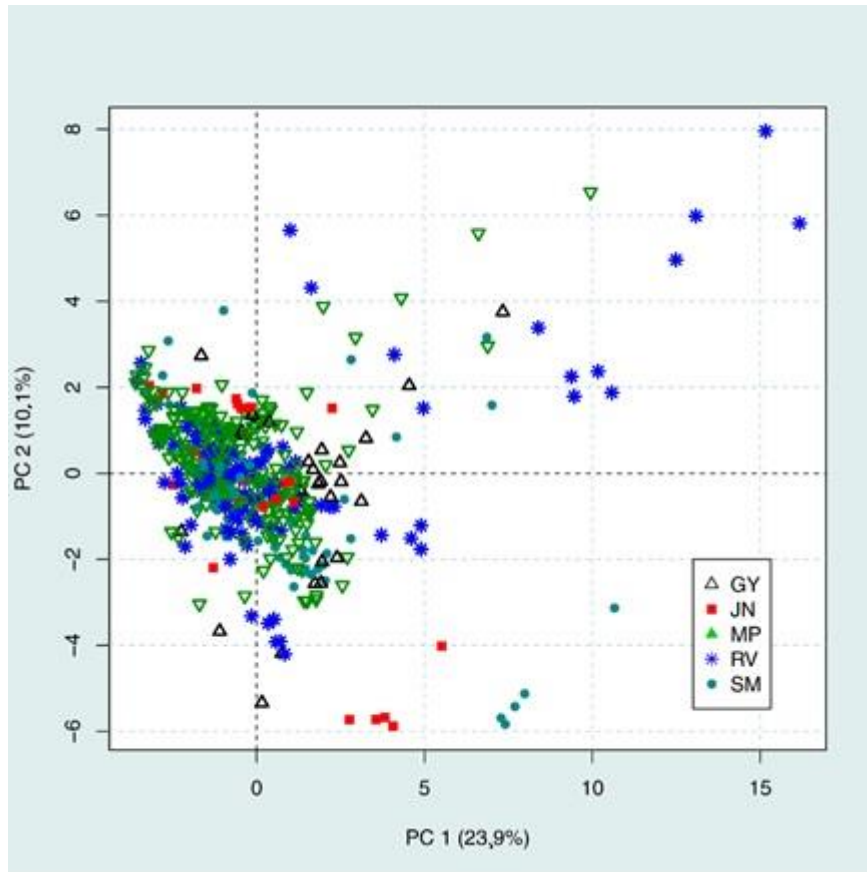
Para poder entender si existen diferencias significativas en los contenidos de los elementos analizados, entre los hollejos de uva de diferentes orígenes geográficos o variedades se empleó la prueba de Kruskal-Wallis. Los resultados revelaron que la mayoría de los parámetros (excepto V, Ag y Te) fueron estadísticamente significativos ($p < 0,05$) entre muestras según su origen geográfico. Por otro lado, la variedad de la uva mostró una menor influencia en el contenido de elementos entre muestras, resultando estadísticamente significativo ($p < 0,05$) solo la varianza del contenido de siete elementos (Fe, Co, Cu, La, Sm, Tm y Lu). La prueba de Kruskal-Wallis mostró que los contenidos minerales pueden considerarse buenos descriptores para construir modelos de clasificación con el objetivo de diferenciar hollejos de uva de acuerdo con su origen geográfico.

4.1.2.3. Análisis PCA

En este trabajo de tesis doctoral, el análisis PCA se realizó con los datos de las concentraciones de los 29 elementos analizados en las muestras de hollejos de uva. Los resultados mostraron que las dos primeras componentes explicaron el 34% de la varianza total. La primer componente principal (PC1) representó el 23,9%, requiriendo casi 12 PC para explicar el 80% de la varianza.

Para visualizar las tendencias de los datos de concentraciones de las 413 muestras de hollejos de uva en un espacio 2D, se presenta la figura 4.5 que muestra los gráficos más importantes de PCA. La Fig. 4.5a presenta el gráfico de puntuación o scores de las dos primeras componentes principales. En el mismo, se puede observar una gran superposición entre las muestras en el centro del eje gráfico, no distinguiéndose una separación de acuerdo al origen geográfico. La figura 4.5b muestra el gráfico de las variables o loadings, que ilustra la influencia de cada variable proyectada en el espacio de las componentes principales. Se observa que la primera PC se asocia fuertemente con los valores de Sm, Gd, Y, Pr, Tm, Nd, U, V y Ga, lo que indica mayores concentraciones de estos elementos en las muestras que mostraron puntuaciones positivas en PC1 en la Fig. 4.5a. Por otro lado, Fe, Ni, Mn, Co, Cu y Zn fueron las variables dominantes en la segunda componente principal. Los scores negativos en PC2 correspondieron a altas concentraciones de estos elementos y los scores positivos a altas concentraciones de Cs y Nb. Los resultados obtenidos por PCA mostraron que existe una gran similitud entre las muestras al considerar las concentraciones de los elementos analizados en hollejos de uva. Por lo tanto, fueron necesarios métodos supervisados avanzados para modelar la clasificación basada en la procedencia.

a)



b)

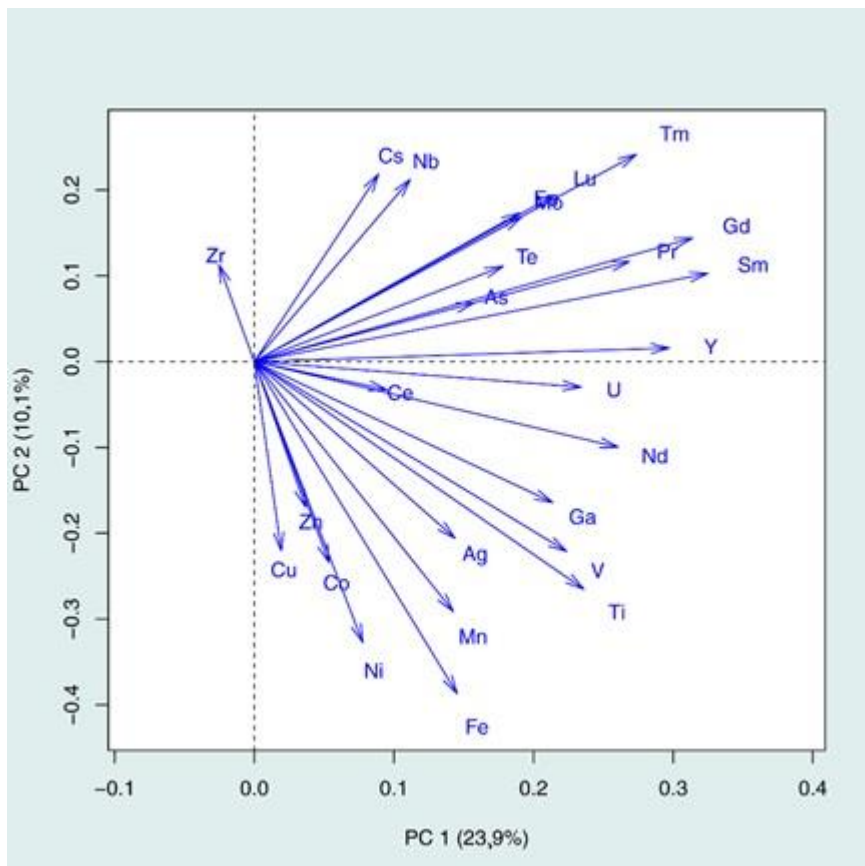


Figura 4.5. Gráfico PCA de a) observaciones y b) variables de la PC1 vs. PC2 de los hollejos.

4.1.2.4. Análisis Predictivo

Una vez realizado el análisis exploratorio de los datos, se procedió a caracterizar las muestras de hollejo de uva en base al perfil mineral mediante técnicas avanzadas de minería de datos para reconocer su origen geográfico.

Los métodos de minería de datos utilizados en este trabajo, incluyeron MLR, SVM, k-NN y RF. Se seleccionaron estos cuatro modelos para evaluar la posible clasificación geográfica de los hollejos de uva, de acuerdo a los métodos que han obtenido buenos resultados de clasificación en trabajos publicados en la literatura de minería de datos.

4.1.2.4.1. Estimación de la exactitud de la clasificación

Los resultados de un método de clasificación aplicado a un conjunto de datos deben ser evaluados para conocer el rendimiento de la clasificación. Para este propósito, es necesaria una separación estricta en los datos en entrenamiento (70%) y prueba (30%), y se aplican los conceptos de optimización de la complejidad de los modelos, utilizando la técnica de validación cruzada (CV) [22]. La matriz de datos se divide aleatoriamente en un conjunto de entrenamiento (70% de los objetos), con miembros de clase conocida utilizados exclusivamente para optimizar los parámetros que son necesarios para cada método; y el conjunto de prueba (30%) que contiene los objetos restantes no incluido en el entrenamiento, utilizado exclusivamente para evaluar el desempeño de cada modelo contra un conjunto de muestras desconocidas para el modelo. En la validación cruzada k-fold, el conjunto de datos se divide aleatoriamente en k subconjuntos mutuamente excluyentes de aproximadamente el mismo tamaño. El clasificador es entrenado y probado k veces, cada vez. La estimación de exactitud de la validación cruzada es el número total de clasificaciones correctas, dividido por el número de instancias en el conjunto de datos. La validación cruzada se repite n veces para cada modelo de clasificación y la estimación final de la exactitud es la media de todas las estimaciones calculadas [23].

Una matriz de confusión construida a partir de los resultados de la predicción contiene información sobre la clasificación real y predicha realizada por un algoritmo de clasificación. Se llama verdadero positivo (TP) y verdadero negativo (TN), respectivamente, al número de muestras positivas y negativas que se clasificaron correctamente. Un falso positivo (FP) se refiere al número de muestras negativas que fueron clasificadas erróneamente como positivas, y un falso negativo (FN) representa el número de instancias positivas que fueron clasificadas erróneamente como negativas [23]. Estos valores se utilizan para estimar las medidas de desempeño del algoritmo, como la tasa de exactitud general, la sensibilidad y la especificidad. La sensibilidad se refiere a muestras pertenecientes a la clase y clasificadas correctamente en esta clase, mientras que la especificidad significa muestras que no pertenecen a la clase modelada y clasificada correctamente como no pertenecientes.

4.1.2.4.2. Optimización de los algoritmos de clasificación

Para clasificar las muestras de hollejos de uva de acuerdo con su procedencia geográfica, se seleccionaron y probaron cuatro modelos de minería de datos que incluyeron MLR, k-NN, SVM y RF. La selección de los parámetros de aprendizaje apropiados es un paso crítico para obtener métodos bien ajustados [22]. Los valores de los parámetros optimizados se calcularon usando una validación cruzada 10 iteraciones repetidas tres veces, teniendo en cuenta la tasa de accuracy o exactitud máxima de clasificación para cada modelo. Como ya se mencionó en la sección anterior, los datos se dividieron aleatoriamente en un conjunto de entrenamiento y prueba usando proporciones 70/30 respectivamente. En el muestreo aleatorio, la división en subgrupos se realizó de manera estratificada, es decir, la proporción de cada clase en la matriz original se mantuvo en los nuevos subgrupos.

Con respecto a los métodos de minería de datos evaluados en este trabajo de tesis doctoral, debe tenerse en cuenta que el modelo MLR no requirió parametrización, mientras que los modelos k-NN, SVM y RF si requirieron optimización de sus parámetros. Para la optimización de k-NN, debido a su simplicidad, sólo una cantidad mínima de tiempo fue

necesaria para generar modelos para cada iteración de la CV. En este método, se necesitó el ajuste de un parámetro, k , correspondiente al número de vecinos a considerar para predecir la pertenencia al grupo de un punto de datos de prueba. En este trabajo, se obtuvo la mejor exactitud promedio al usar $k = 5$.

En términos de modelos más complejos se encuentran SVM y RF, para los cuales se necesita más tiempo para lograr su parametrización. La optimización de parámetros más importante para SVM incluye la especificación de los parámetros de la función Kernel y el parámetro de costo C . Se seleccionó la función de base radial (RBF) para la función Kernel del algoritmo. Esta función Kernel es rápida para calcular los parámetros óptimos del modelo. La identificación de una combinación "óptima" de valores de parámetros fue un desafío. El método de búsqueda Grid se usó para determinar los parámetros óptimos al intentar diferentes valores y seleccionar aquellos que tienen el menor valor de error de entrenamiento. Esto significa que, para que cada parámetro se optimice, se prescribe un número finito de valores posibles, y luego se consideran todas las combinaciones posibles de estos valores hasta que se encuentra una que produce el mejor resultado [24]. La figura 4.6 muestra los valores óptimos encontrados para los parámetros C y γ , donde muestra que los parámetros $C = 16$ y $\gamma = 0,036$ fueron los mejores para obtener el error de rendimiento mínimo en la configuración de entrenamiento por 10 iteraciones de CV.

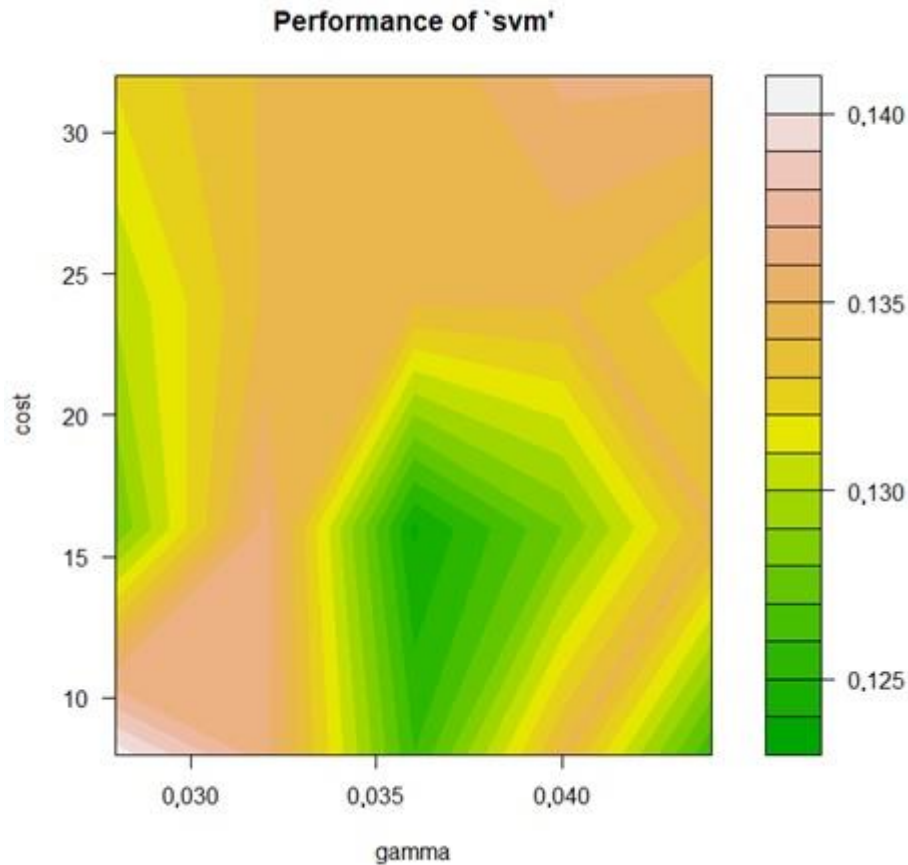


Figura 4.6. Optimización de SVM.

El algoritmo RF contó con dos parámetros a optimizar, *mtry* y *nt*, correspondientes a número total de variables a considerar para dividir en cada nodo (muestreo estratificado) y al número de árboles, respectivamente. Inicialmente se utilizó la optimización por defecto, donde no se utilizó el re muestreo previo al procesamiento con validación cruzada (10 iteraciones, repetido 3 veces) y el parámetro de ajuste 'mtry' se mantuvo constante en un valor de 5.38, correspondiente a la raíz cuadrada del número total de variables disponibles. Para el número de árboles, el valor predeterminado fue de 500. Luego de realizar la optimización por defecto, con la cual se logró un buen rendimiento del algoritmo resultante; se buscó mejorar la performance del modelo mediante la optimización Grid, con la cual se obtuvo un valor optimizado de mtry de 24 y nt seleccionado de 1500. En la Fig. 4.7 se puede observar que con un número de árboles de 1500, la accuracy y el valor kappa del modelo presentaron el mayor valor. El estadístico Kappa varía en el intervalo entre 0 y 1, cuanto mayor

sea el valor de Kappa significa mayor acuerdo / vinculación entre dos conjuntos de datos categorizados [25]. En la Fig. 4.8 se presenta una comparación de los dos métodos de optimización de RF para el valor de mtry, se observa un mayor valor de accuracy con la optimización Grid con 24 mtry.

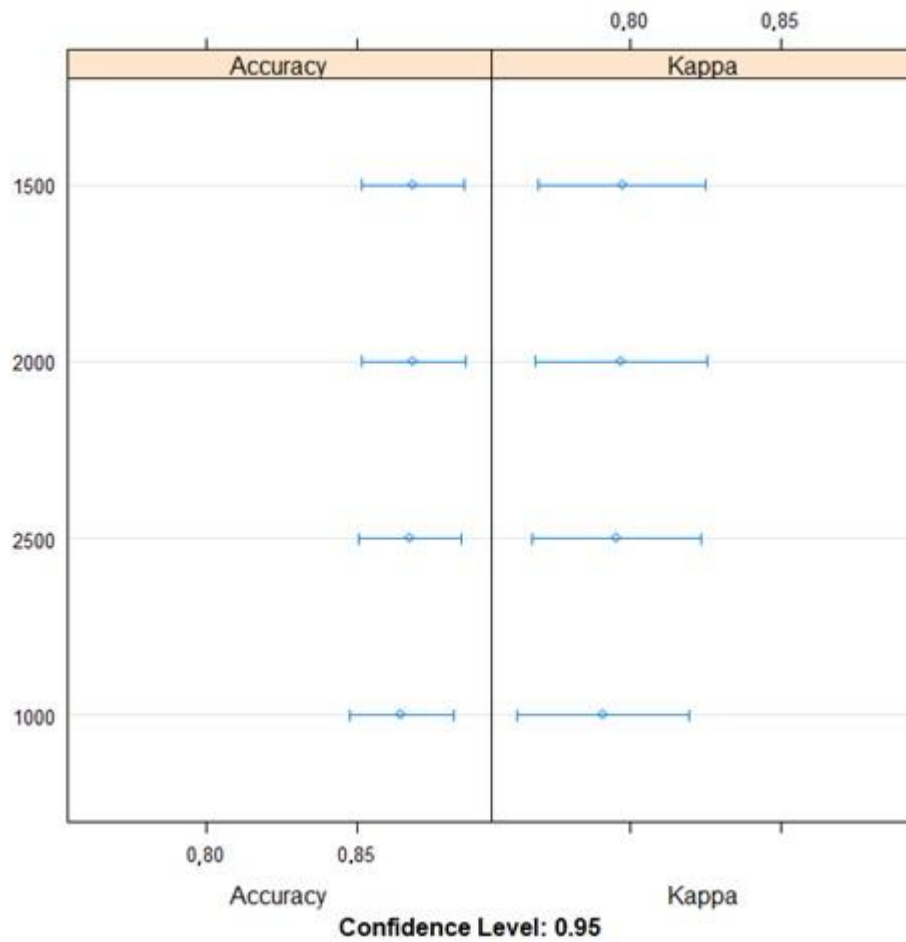
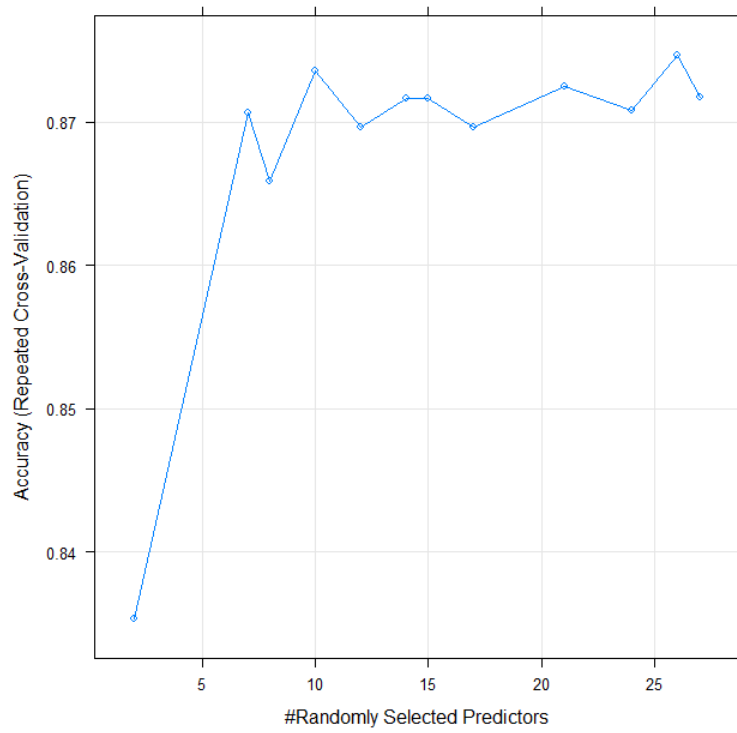


Figura 4.7. Número óptimo de árboles de clasificación para RF.

a)



b)

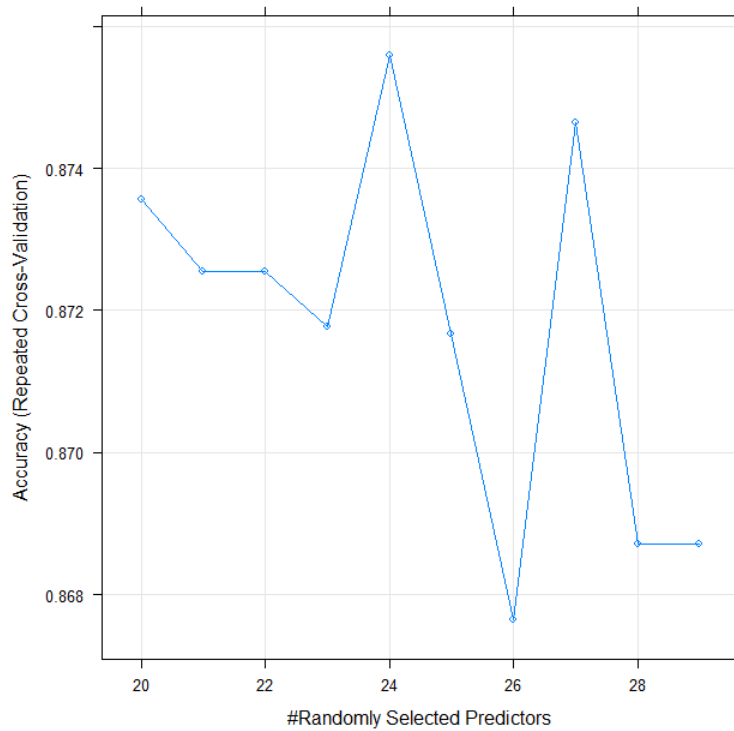


Figura 4.8. Comparación de la optimización de mtry de RF por defecto (a) y grid (b).

4.1.2.4.3. Clasificación geográfica de hollejos de uva de la provincia de Mendoza

Una vez seleccionados los valores óptimos de los parámetros para cada algoritmo, la sensibilidad, la especificidad y la tasa de accuracy general de los modelos de minería de datos aplicados se consideraron para evaluar el rendimiento de la clasificación en el conjunto de muestras de prueba o testeo. Los resultados indicando el rendimiento de los diferentes métodos de clasificación se muestran en la Tabla 4.4.

Tabla 4.4. Comparación de la performance de los diferentes métodos de minería de datos aplicados en las muestras de prueba.

Grupos	MLR		k-NN (k = 5)		SVM (C = 16; $\epsilon = 0,036$)		RF (nt = 500; mtry = 5)		RF-Grid (nt = 1500; mtry = 24)	
	Sens.	Espec.	Sens.	Espec.	Sens.	Espec.	Sens.	Espec.	Sens.	Espec.
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
GY	88	100	50	100	87	100	87	100	87	99
JN	66	80	12	98	75	98	87	100	87	98
MP	80	87	94	86	91	85	99	88	100	89
RV	64	66	78	91	75	94	75	97	72	96
SM	60	37	64	92	76	97	80	97	76	98
Accuracy (%)	74,0		78,5		84,0		88,9		88,2	

A partir del análisis de la Tabla 4.4 se observa que los cuatro métodos probados mostraron diferente grado de éxito en la predicción de las muestras de prueba, aunque dentro de un rango pequeño. El orden de las tasas de clasificación exitosas fue el siguiente: RF> RF-grid> SVM> k-NN> MLR. El único modelo lineal probado, MLR, presentó el peor desempeño desde el punto de vista del valor de accuracy. El uso de métodos no lineales fueron los más efectivos para resolver el problema de clasificación geográfica de hollejos de uva. El modelo RF con optimización por defecto fue el mejor modelo para discriminar las muestras de hollejos según su procedencia, con una exactitud de clasificación de casi el 90%, mientras que el mismo algoritmo con optimización grid logró la segunda mejor exactitud de clasificación. Estos resultados también se observaron en el rendimiento de clasificación del conjunto de datos de entrenamiento presentado en la Tabla 4.5, la cual muestra los valores de accuracy comparativos para muestras de entrenamiento y prueba. En conclusión, se obtuvo una buena discriminación de las muestras en estudio con estos modelos no lineales, lo que puede explicarse por la flexibilidad y la capacidad del algoritmo para crear un modelo generalizado, incluso para un pequeño número de muestras, lo que también se puede encontrar en los buenos resultados obtenidos en la reciente literatura con aplicación de técnicas de minería de datos.

Tabla 4.5. Resultados de la clasificación de los diferentes modelos aplicados a los conjuntos de entrenamiento y prueba.

Algoritmos	Total de muestras		Accuracy (%)	
	Entrenamiento	Prueba	Entrenamiento	Prueba
MLR	290	123	86,2	74
k-NN	290	123	80,7	78,5
SVM	290	123	81,7	84
RF	290	123	90,8	88,9
RF-Grid	290	123	89	88,2

Una vez obtenido el mejor método de clasificación con RF, se procedió a calcular la importancia de las variables (Random Forest Importance, RFI). RF ofrece la posibilidad de obtener un ranking de los predictores, lo cual refleja la importancia de las variables en la predicción de la respuesta y, por lo tanto, la oportunidad de poder elegir las variables con la mejor capacidad predictiva. La Fig. 4.9 presenta las 10 variables más importantes (Rb, Mn, Zn, Cu, Ni, Ce, Tl, As, Ti y La) para el método de clasificación de hollejo de uva aplicando RF. Rb tiene el mayor poder de discriminación, con una importancia del 100% para la clasificación geográfica del hollejo, seguido de Mn, Zn en importancia. El algoritmo de RF utiliza el método de out-of-bag (OOB) o muestras fuera de bolsa en su traducción al español, para medir la accuracy y el puntaje de cada variable en el ranking, cuantificando el impacto de las mismas en la exactitud de la predicción. El procedimiento se basa en una prueba de permutación. La prueba de permutación testea que la variable es independiente de la respuesta y de todos los demás predictores de la clasificación. Los puntajes reflejan la contribución que cada predictor hace en la clasificación de la variable objetivo. Con el uso de esta función del algoritmo RF, se puede llevar a cabo la clasificación con sólo el mejor subconjunto de variables, lo cual es ventajoso para un futuro estudio del mismo tipo de muestras.

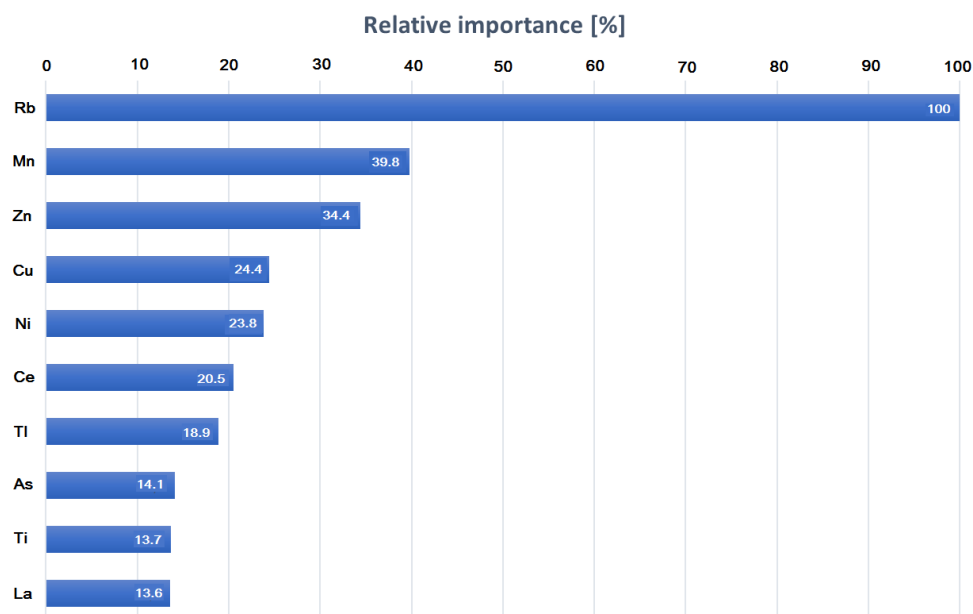


Figura 4.9. Importancia de las variables en RF.

4.1.2.5. Conclusiones

En conclusión, los resultados mostraron que a través de la aplicación de cuatro algoritmos de minería de datos combinados con un análisis multielemental mediante ICP-MS, se logró la clasificación geográfica de muestras de hollejos de uva procedentes de cinco regiones vitivinícolas cercanas. El ranking de la importancia de las variables, nos permitió encontrar qué elementos fueron relevantes para la discriminación de las muestras.

El método de RF mostró supremacía en la predicción del origen geográfico. SVM presentó también un buen desempeño. Sin embargo, la optimización de parámetros para RF mostró menos variación en los resultados y fue más consistente (promedio de la accuracy de predicción) que los otros algoritmos. Además, RF convergió más rápido a soluciones finales que SVM, lo que indicó la necesidad de un costo computacional más bajo.

Los resultados informados pueden ser útiles para fines de autenticación de uvas destinadas a la producción de vino. Por ejemplo, los productores pueden contar con un control de calidad de las materias primas utilizadas en la industria vitivinícola. Comprender el comportamiento de los componentes químicos también es importante, ya que sirve como fuente de información para la industria del vino para garantizar la calidad de las uvas y por ende de los vinos, lo cual es un aspecto relevante en la región, ya que el vino es uno de los principales productos en la provincia de Mendoza, Argentina.

4.1.2.6. Referencias

- [1] International Organization of Vine and Wine (OIV), State of the sector in 2018, 2019. <http://www.oiv.int/public/medias/6679/en-oiv-state-of-the-vitiviniculture-world-market-2019.pdf>.
- [2] Instituto Nacional de Vitivinicultura (INV), Informe anual de cosecha y elaboración 2018, (2018). www.inv.gob.ar (accessed October 10, 2019).
- [3] M. Likar, K. Vogel-Mikuš, M. Potisek, K. Hančević, T. Radić, M. Nečemer, M. Regvar, Importance of soil and vineyard management in the determination of grapevine

- mineral composition, *Sci. Total Environ.* 505 (2015) 724–731. doi:10.1016/j.scitotenv.2014.10.057.
- [4] M. Tanner, D. Günther, Short transient signals, a challenge for inductively coupled plasma mass spectrometry, a review, *Anal. Chim. Acta.* 633 (2009) 19–28. doi:10.1016/j.aca.2008.11.041.
- [5] B. Heung, H.C. Ho, J. Zhang, A. Knudby, C.E. Bulmer, M.G. Schmidt, An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping, *Geoderma.* 265 (2016) 62–77. doi:10.1016/j.geoderma.2015.11.014.
- [6] L. Mutihac, R. Mutihac, Mining in chemometrics, *Anal. Chim. Acta.* 612 (2008) 1–18. doi:10.1016/j.aca.2008.02.025.
- [7] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 2016.
- [8] N. Mastrogiannis, B. Boutsinas, I. Giannikos, A method for improving the accuracy of data mining classification algorithms, *Comput. Oper. Res.* 36 (2009) 2829–2839. doi:10.1016/j.cor.2008.12.011.
- [9] E. Raczko, B. Zagajewski, Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images, *Eur. J. Remote Sens.* 50 (2017) 144–154. doi:10.1080/22797254.2017.1299557.
- [10] D. Bertoldi, R. Larcher, M. Bertamini, S. Otto, G. Concheri, G. Nicolini, Accumulation and distribution pattern of macro- and microelements and trace elements in *Vitis vinifera* L. cv. Chardonnay berries., *J. Agric. Food Chem.* 59 (2011) 7224–7236. doi:10.1021/jf2006003.
- [11] A. Cugnetto, L. Santagostini, L. Rolle, S. Guidoni, V. Gerbi, V. Novello, Tracing the “terroirs” via the elemental composition of leaves, grapes and derived wines in cv Nebbiolo (*Vitis vinifera* L.), *Sci. Hortic. (Amsterdam).* 172 (2014) 101–108. doi:10.1016/j.scienta.2014.03.055.

- [12] M. De Nisco, M. Manfra, A. Bolognese, A. Sofo, A. Scopa, G.C. Tenore, F. Pagano, C. Milite, M.T. Russo, Nutraceutical properties and polyphenolic profile of berry skin and wine of *Vitis vinifera* L. (cv. Aglianico), *Food Chem.* 140 (2013) 623–629. doi:10.1016/j.foodchem.2012.10.123.
- [13] S.S. Mitić, M. V. Obradović, M.N. Mitić, D.A. Kostić, A.N. Pavlović, S.B. Tošić, M.D. Stojković, Elemental Composition of Various Sour Cherry and Table Grape Cultivars Using Inductively Coupled Plasma Atomic Emission Spectrometry Method (ICP-OES), *Food Anal. Methods.* 5 (2012) 279–286. doi:10.1007/s12161-011-9232-2.
- [14] S.Y. Rogiers, D.H. Greer, J.M. Hatfield, B.A. Orchard, M. Keller, Mineral sinks within ripening grape berries (*Vitis vinifera* L.), *Vitis - J. Grapevine Res.* 45 (2006) 115–123.
- [15] E.C. Sousa, A.M.A. Uchôa-Thomaz, J.O.B. Carioca, S.M. De Moraes, A. De Lima, C.G. Martins, C.D. Alexandrino, P.A.T. Ferreira, A.L.M. Rodrigues, S.P. Rodrigues, J. do N. Silva, L.L. Rodrigues, Chemical composition and bioactive compounds of grape pomace (*Vitis vinifera* L.), Benitaka variety, grown in the semiarid region of Northeast Brazil, *Food Sci. Technol.* 34 (2014) 135–142. doi:10.1590/S0101-20612014000100020.
- [16] Y. Yang, C. Duan, H. Du, J. Tian, Q. Pan, Trace element and rare earth element profiles in berry tissues of three grape cultivars, 61 (2010) 401–407. <https://www.scopus.com/record/display.uri?eid=2-s2.0-77956935664&origin=resultslist&sort=plf-f&src=s&st1=Mineral+of++grape+skins&st2=&sid=5aac4a31e2c78600d5fafc9f560459b0&sot=b&sdt=b&sl=38&s=TITLE-ABS-KEY%28Mineral+of++grape+skins%29&relpos=12&citeCnt=15&searchTerm=> (accessed December 28, 2017).
- [17] D. Bertoldi, R. Larcher, G. Nicolini, M. Bertamini, G. Concheri, Distribution of rare earth elements in *Vitis vinifera* L. “Chardonnay” berries, *Vitis.* 48 (2009) 49–51.
- [18] S.S. Mitić, M. V. Obradović, M.N. Mitić, D.A. Kostić, A.N. Pavlović, S.B. Tošić, M.D.

- Stojković, Elemental Composition of Various Sour Cherry and Table Grape Cultivars Using Inductively Coupled Plasma Atomic Emission Spectrometry Method (ICP-OES), *Food Anal. Methods*. 5 (2012) 279–286. doi:10.1007/s12161-011-9232-2.
- [19] J. Ribéreau-Gayon, Y. Glories, A. Maujean, D. Dubourdieu, *The Chemistry of Wine: Stabilization and Treatments*, 2nd ed., John Wiley & Sons, Ltd, France, 2006.
- [20] Secretaría de Estado de Salud Pública, Código Alimentario Argentino. Ley 18284., Artículos 155 Al 183 - Los Prod. Aliment. (2017). <https://www.argentina.gob.ar/anmat/codigoalimentario> (accessed March 4, 2019).
- [21] Official Journal of the European Communities, Commission of the European Communities: Commission Regulation (EC) No.221/2002 of the 6 February 2002 amending regulation (EC) No. 466/2002 setting maximum levels for certain contaminants in foodstuffs., (2002).
- [22] K. Varmuza, P. Filzmoser, Classification, in: *Introd. to Multivar. Stat. Anal. Chemom.*, CRC Press, 2009: pp. 195–204.
- [23] C. Maione, E.S. de Paula, M. Gallimberti, B.L. Batista, A.D. Campiglia, F.B. Jr, R.M. Barbosa, Comparative study of data mining techniques for the authentication of organic grape juice based on ICP-MS analysis, *Expert Syst. Appl.* 49 (2016) 60–73. doi:10.1016/j.eswa.2015.11.024.
- [24] T. Eitrich, B. Lang, Efficient optimization of support vector machine learning parameters for unbalanced datasets, *J. Comput. Appl. Math.* 196 (2006) 425–436. doi:10.1016/j.cam.2005.09.009.
- [25] Y. Kumar, G. Sahoo, Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA, *Int. J. Inf. Technol. Comput. Sci.* 4 (2012) 43–49. doi:10.5815/ijitcs.2012.07.06.

4.2. CLASIFICACIÓN GEOGRÁFICA DE JUGOS DE UVA

4.2.1. Identificación de la procedencia de jugos de uva comerciales de Argentina y Brasil basados en la composición elemental y técnicas quimiométricas

4.2.1.1. Introducción

El jugo de uva se define como la bebida obtenida a partir de la fracción comestible de la uva (*Vitis ssp.*) a través de un proceso tecnológico adecuado. Actualmente, debido a la tendencia a un estilo de vida natural y saludable, su consumo a nivel mundial ha aumentado significativamente [1,2]. Esta bebida es una excelente fuente de una amplia gama de componentes beneficiosos, como nutrientes y compuestos bioactivos, como por ejemplo, carbohidratos, vitaminas, minerales, polifenoles, flavonoides, entre otros [3–5]. En los últimos años, Argentina y Brasil se han convertido en importantes productores de esta bebida, siendo Brasil también un buen consumidor de este jugo. Sin embargo, en el caso de Argentina, la producción de jugo de uva se destina principalmente a la exportación como jugo de uva concentrado [6].

Con respecto al comportamiento de los consumidores, sus demandas en la calidad y autenticidad de alimentos y bebidas han aumentado significativamente. Actualmente, la preocupación se centra no solo en la seguridad de los productos, sino también en la protección contra la adulteración con materias primas de baja calidad y en el etiquetado correcto con la composición, propiedades y origen del producto detallados. Como resultado de esto, se ha desarrollado un sistema de monitoreo preciso para garantizar la calidad y autenticidad de los productos, proporcionando una protección real para los consumidores y a su vez productores confiables [7–10]. El sistema de regulación de la denominación de origen protegida o controlada (DOC) garantiza que la calidad está estrechamente vinculada a su origen territorial, asegurando un conocimiento de la trazabilidad del producto final [7].

El análisis de diversos parámetros químicos de las muestras, es una excelente herramienta para evaluar su origen, presencia de fraude y adulteración. Muchos estudios han

demostrado que los compuestos químicos de las uvas y sus productos derivados, principalmente vinos y jugos, están influenciados por las condiciones climáticas y agrícolas relacionadas con el cultivo de la planta y el crecimiento, y por factores ambientales como la composición del suelo, la composición del agua de riego, exposición a la luz solar, entre otros [8,11,12]. La composición mineral refleja la geoquímica en la que se cultivan las vides y, por lo tanto, los productos derivados de la uva están influenciados por la composición del suelo [13]. El análisis de las concentraciones de elementos en muestras de jugo de uva combinado con herramientas quimiométricas es una excelente manera de identificar el origen geográfico de esta bebida. Existen numerosas técnicas para el análisis elemental, entre ellas, los métodos espectroquímicos como la espectrometría de absorción atómica con llamas (FAAS), la espectrometría de emisión óptica de plasma acoplado inductivamente (ICP-OES) o la espectrometría de masas (ICP-MS), son los más utilizados [14]. ICP-MS ofrece la posibilidad de analizar la concentración multielemental de muestras a niveles de trazas y ultra trazas con una detección de alta sensibilidad [15].

El tratamiento estadístico es indispensable para extraer información de los datos generados y para identificar similitudes y diferencias entre las muestras. La aplicación de métodos quimiométricos ha demostrado ser una herramienta versátil y valiosa para evaluar la autenticidad de alimentos y bebidas [9]. En quimiometría, los métodos de reconocimiento de patrones son poderosos para determinar el origen geográfico de los productos alimenticios. Estas técnicas se dividen en métodos supervisados y no supervisados. Los métodos no supervisados no requieren ningún conocimiento a priori sobre la configuración grupal de los datos, este tipo de análisis es muy útil como enfoque exploratorio en la etapa inicial de una investigación y puede presentarse con una representación visual simple, como el gráfico de análisis de componentes principales (PCA) o un dendograma en el análisis de agrupamiento jerárquico (HCA). Estas herramientas son útiles a la hora de controlar la autenticidad de

productos, asegurando el origen de la región de producción e identificando cualquier posible adulteración del mismo [4,16].

El objetivo de este trabajo de tesis doctoral fue evaluar la posibilidad de distinguir el origen geográfico de jugos de uva de dos importantes regiones productoras de Sudamérica, como lo son Brasil y Argentina, en base a un análisis de múltiples elementos por la técnica ICP-MS, combinada principalmente con la implementación de herramientas simples de reconocimiento de patrones no supervisadas.

4.2.1.2. Concentración multielemental de los jugos de uva

En la Tabla 4.6 se presentan las concentraciones de los elementos analizados en las muestras de jugo de uva. En ella se muestran los valores promedios y de desviación estándar expresados en $\mu\text{g/L}$ de cada elemento determinado en función de su país de origen.

Tabla 4.6. Media y desviación estándar de la concentración de los elementos analizados en muestras de jugo de uva según su país de origen.

Elemento	LOD	Argentina		Brasil	
		(n = 16)		(n = 10)	
		Media	DE	Media	DE
As	0,01	28,2	47,8	6,7	9,6
Cr	0,2	41,3	7,6	21,5	16,5
Cu	0,1	515	408	297	223
Fe	2,9	2409	1315	1037	685
Mn	0,03	680	543	1041	678
Mo	0,01	12,3	16,9	4,8	4,9
Ni	0,05	20,6	9,7	31,3	17,9
Pd	0,06	4,1	2,3	2,9	1,0
Rb	0,06	417	338	1794	1373
V	0,05	24,6	33,3	15,8	22,5
Y	0,003	0,11	0,10	0,01	0,03

Del análisis de los resultados presentados en la Tabla 4.6, se observa que Fe fue el elemento más abundante encontrado en las muestras de jugo de uva. El nivel medio más alto de este elemento se encontró en las muestras de Argentina (2409 µg/L) con más del doble de concentración del encontrado en Brasil (1037 µg/L). Rb fue el elemento más alto encontrado en los jugos de uva de Brasil; este elemento presentó una concentración media de 1794 µg/L, mientras que en los jugos de Argentina este elemento se cuantificó con una concentración media de 417 µg/L. Los siguientes elementos que siguieron en orden de concentración fueron Mn y Cu. La concentración más alta de Mn se encontró en muestras de jugo de uva de Brasil (1041 µg/L), mientras que la concentración media de este elemento en muestras de Argentina fue de 680 µg/L. Con respecto al elemento Cu, se observó que en las muestras de Argentina la concentración media fue de 515 µg/L y en los jugos de Brasil, este elemento obtuvo una concentración media de 297 µg/L. Los niveles de concentración de los elementos Fe, Mn y Cu en los jugos de uva analizados, estuvieron de acuerdo con los reportados por otros autores [14,17]. Con respecto a otros trabajos sobre jugos de uva publicados, se encontraron diferencias en los niveles de estos elementos [2,18–23]. En concentraciones naturales, estos elementos participan en importantes funciones biológicas para el organismo humano, por lo que resultan ser elementos esenciales para la vida. En el caso de Rb, las concentraciones estuvieron de acuerdo con lo reportado en trabajos publicados previamente [4,17,22].

Los siguientes elementos cuantificados en niveles menores de concentración fueron Cr, Ni, As, V y Mo, con concentraciones medias que estuvieron de acuerdo con trabajos publicados previamente [14,17,20]. El Cr es un micronutriente esencial para animales y plantas pero en concentraciones elevadas se considera como un elemento nocivo, como el As, el cual se sabe que es un elemento tóxico para la vida humana. Finalmente, Pd e Y se encontraron en baja concentración o no se detectaron en las muestras de jugos de uva.

Para hacer visibles las diferencias en las concentraciones de los elementos analizados en las muestras de jugo de los dos países en estudio, se presenta la Fig. 4.10. En esta figura, se

pueden comparar las gráficas de caja y bigote del rango y distribución de las concentraciones de elementos analizados en las muestras de jugo de uva de Argentina (AR) y Brasil (BR). Se pueden observar claras diferencias en las concentraciones de Cr, Fe, Pd y Rb entre los dos países, lo que indica que podrían ser variables de interés para el problema de clasificación geográfica.

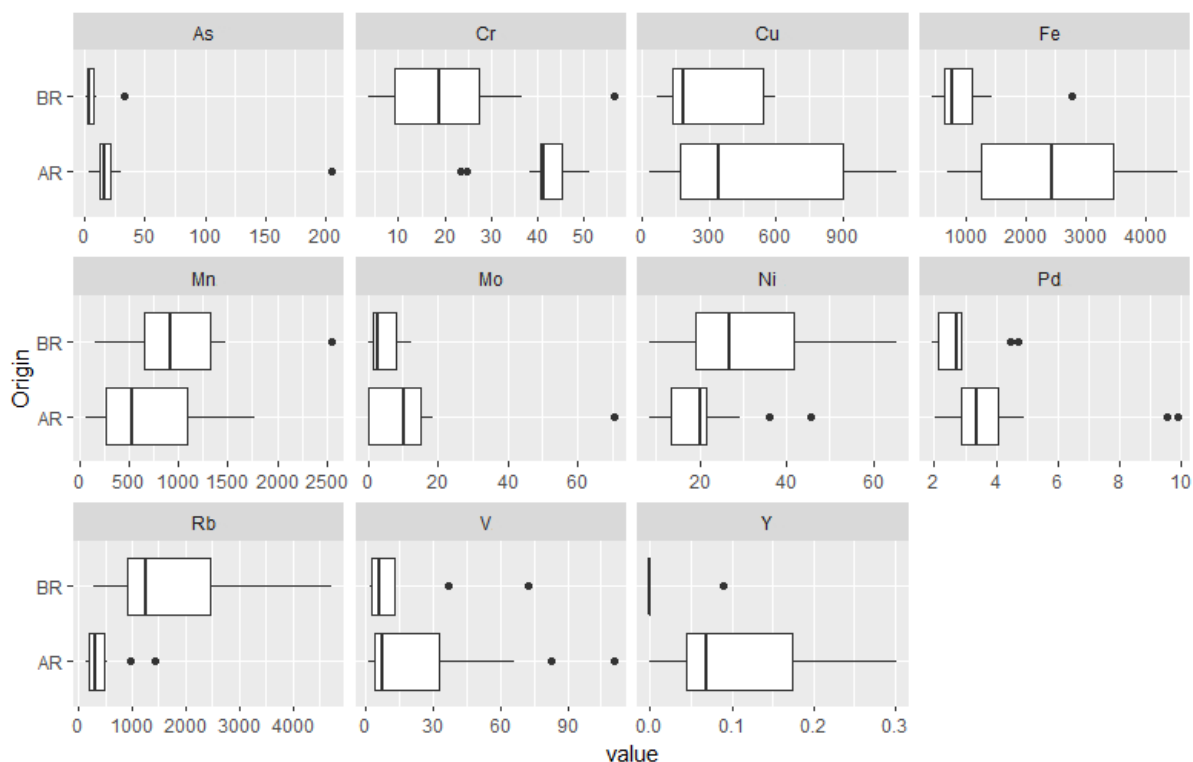


Figura 4.10. Gráficos box-plot de los once elementos determinados en los jugos de uva de acuerdo al país de procedencia.

Finalmente, se estudiaron las relaciones lineales entre las concentraciones de los elementos analizados en los jugos de uva y para ello, se utilizó el coeficiente de correlación de Pearson (r). En la Fig. 4.11 se presenta un gráfico con el diferente grado de correlación entre todas las variables. Se puede observar una alta correlación positiva entre Mo y As ($r=0,94$), es decir, al aumentar la concentración de uno aumenta el del otro, también se observó una alta correlación positiva ($r = 0,7$) para los pares de elementos Rb-Mn. Para el resto de las variables medidas las correlaciones fueron moderadas o casi insignificantes. Este análisis nos permite encontrar aquellas variables que pueden ser descartadas a la hora del problema de

clasificación geográfica, debido a su alta correlación y por ende comportamiento similar con otras variables.

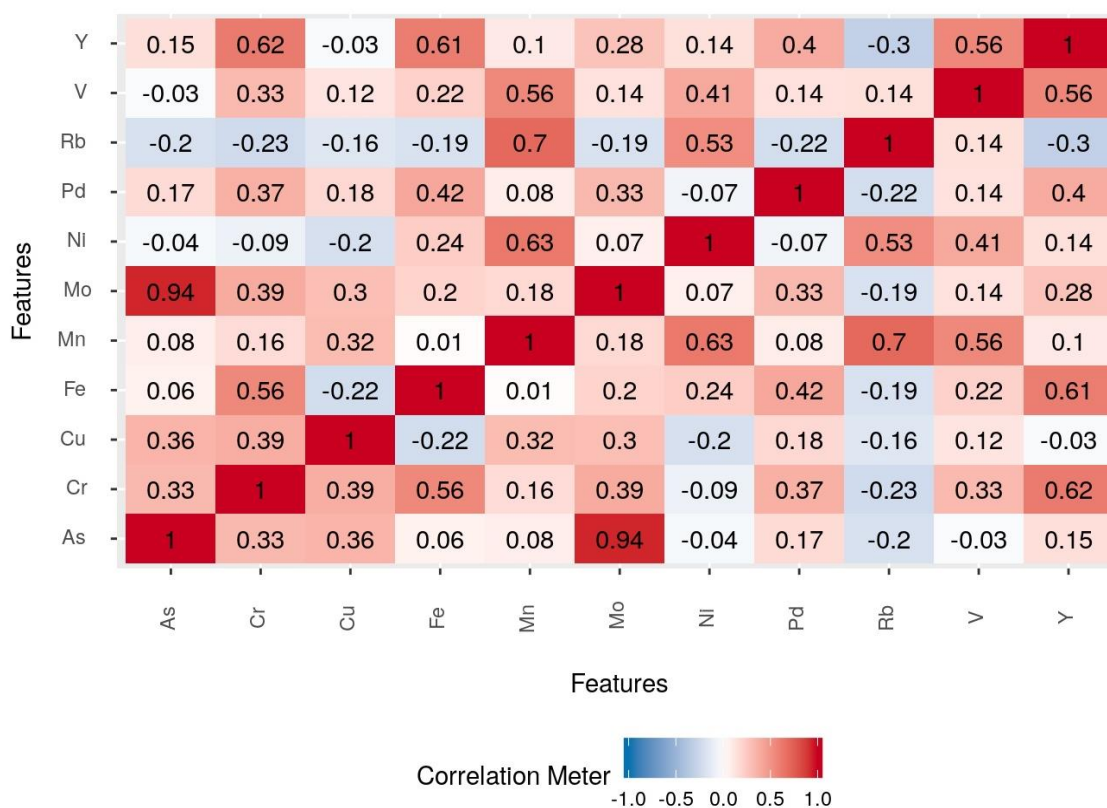


Figura 4.11. Coeficientes de correlación de Pearson (r).

4.2.1.3. Análisis exploratorio

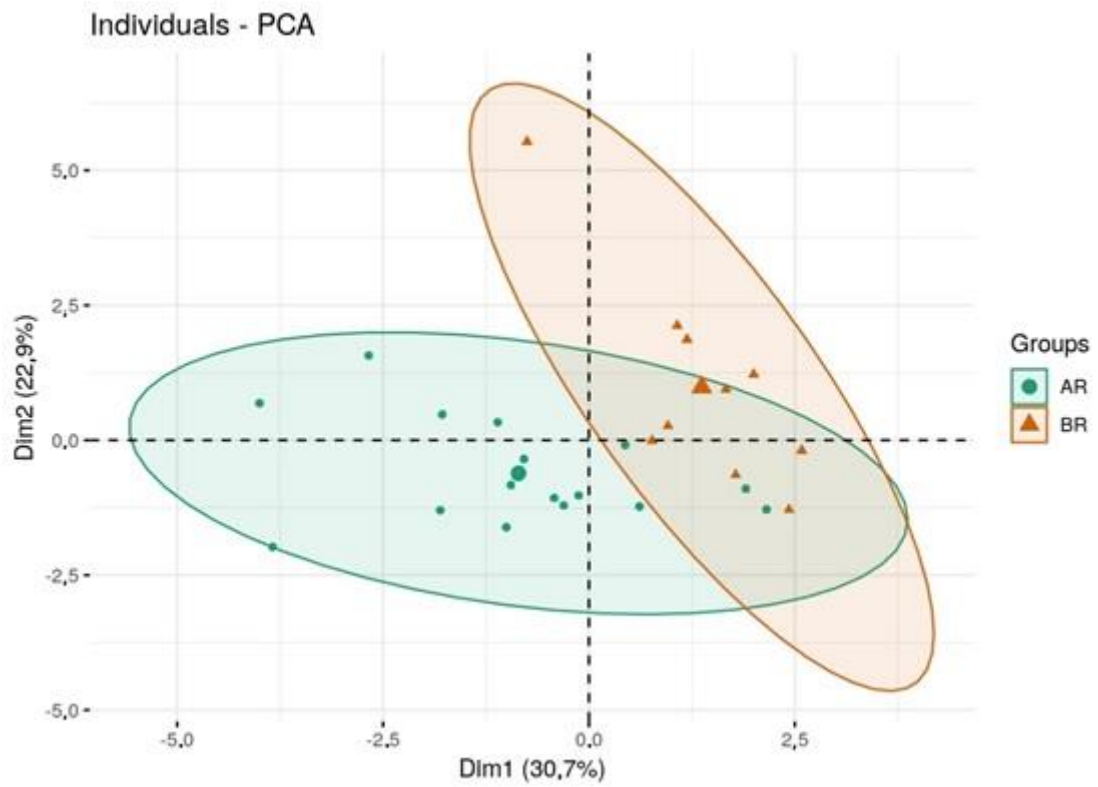
Inicialmente se llevó a cabo la caracterización quimiométrica básica de las muestras de jugo de uva estudiadas mediante la técnica de PCA. Como se indicó anteriormente, PCA es un método analítico quimiométrico que reduce las dimensiones de los datos para hacer posible la visualización de la agrupación natural de los objetos estudiados (muestras de jugo de uva) en el plano o espacio 3D de los componentes principales más importantes. En este trabajo de tesis doctoral, se utilizaron los datos de concentración de los 11 elementos para realizar el análisis de PCA.

El contenido total de la información dada por las componentes principales se expresó por como un valor porcentual acumulado de la varianza total. Los resultados mostraron que las dos primeras componentes principales explicaron el 53,6% de la varianza total. La primer

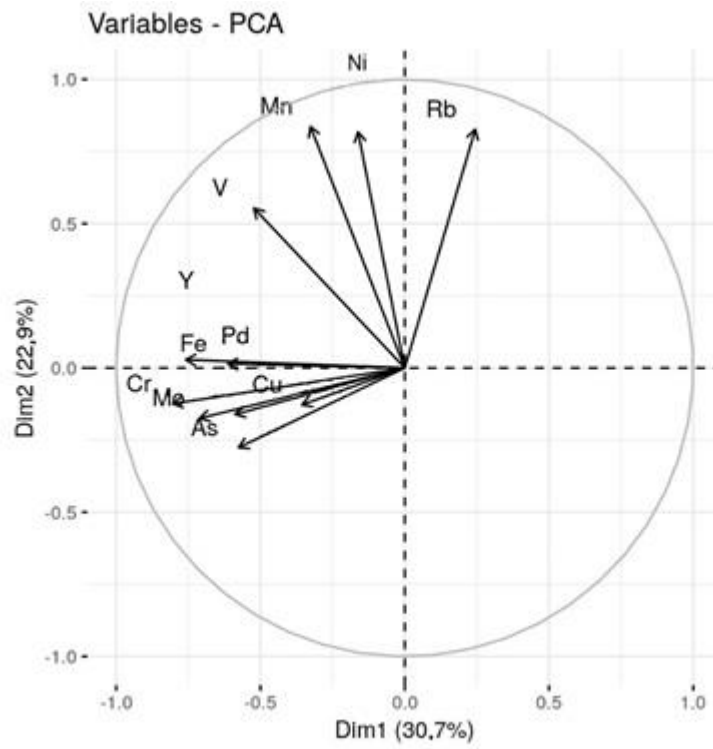
componente principal (PC1) representó el 30,7%, y la siguiente componente principal (PC2), el 22,9%, representando las tres primeras componentes principales el 70,3% de la variabilidad total de los datos. En la figura 4.12 se presentan los gráficos de PCA más importantes, el gráfico de puntuación o scores (figura 4.12a) y el gráfico del peso de las variables o loadings (figura 4.12b) de las dos primeras componentes principales. En el gráfico de scores se puede observar una buena separación entre las muestras de jugos de Argentina y Brasil. Las muestras de AR mostraron puntuaciones principalmente negativas en la PC1 y se diferenciaron de las muestras de BR, las cuales presentaron scores positivos en la PC1.

La figura 4.12b muestra la importancia de cada variable para las componentes principales más informativas PC1 y PC2. La PC1 se asoció fuertemente con los valores de Cr, Y, Mo, Fe, Pd y As, lo que indica concentraciones más altas en muestras que mostraron puntuaciones negativas (muestras de AR). Por otro lado, Mn, Rb, Ni y V fueron las variables dominantes en la segunda componente principal, y las muestras con puntajes positivos en PC2 (jugos de BR) correspondieron a altas concentraciones de estos elementos. Para facilitar la visualización de la importancia relativa de cada elemento o variable en las componentes principales, se presenta la Fig. 4.12c. Los resultados obtenidos mediante el análisis de PCA mostraron que a partir del análisis de 11 elementos, las muestras de jugo de uva pudieron ser diferenciadas en términos de su país de origen.

a)



b)



c)

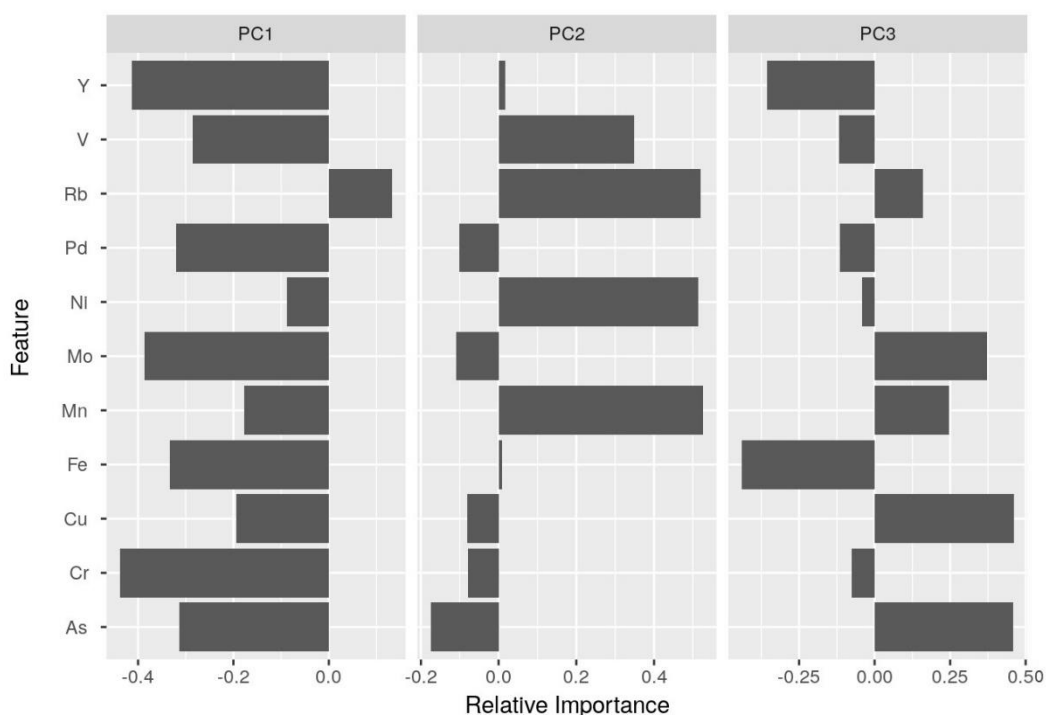


Figura 4.12. Gráficos de PCA de las observaciones (a) y variables (b) de la PC1 versus PC2.

Importancia relativa de los elementos para las tres primeras componentes principales (c).

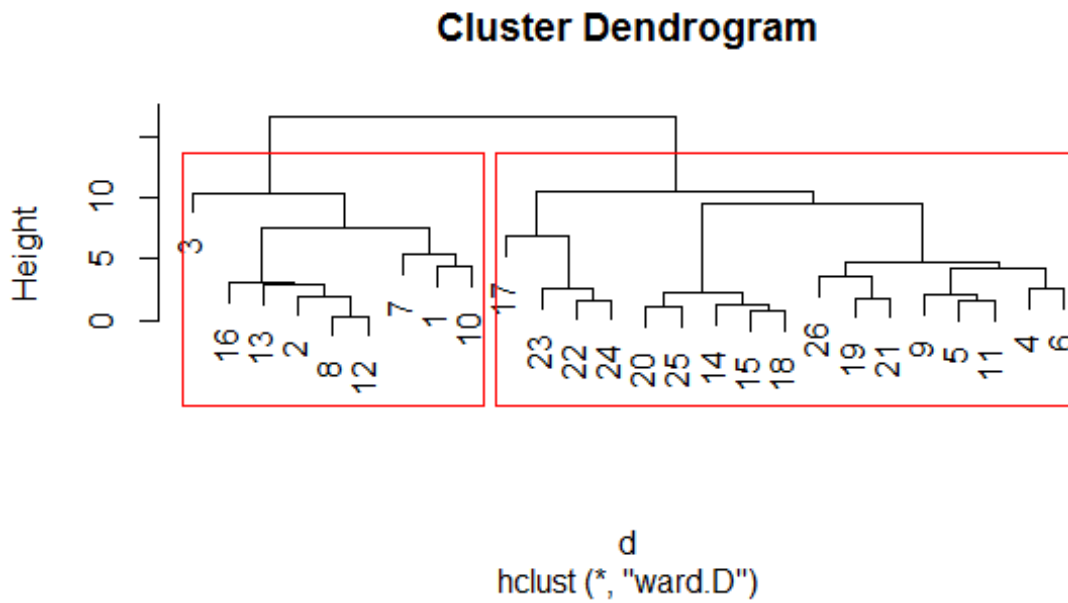
Luego de aplicar PCA a los datos de concentración obtenidos, se evaluó otra técnica exploratoria no supervisada como el análisis de clúster o conglomerados (CA), para caracterizar el papel de los elementos analizados en la clasificación de los jugos de uva en función de su país de origen. En este trabajo, se utilizó el análisis de conglomerados jerárquico (HCA, por sus siglas en inglés) para agrupar muestras experimentales en conglomerados, en función de la similitud dentro de la clase y la diferencia entre grupos, según el país de origen. Se empleó el método de Ward con el enfoque clásico para medir la distancia, es decir distancia Euclidiana. En la figura 4.13a se presenta el dendograma de HCA aplicado a los datos para observar las características de la agrupación de las muestras. Al analizar el dendograma, es posible identificar dos grandes grupos. El primero estuvo compuesto por nueve muestras, de las cuales todas las muestras correspondieron a jugos de Argentina. El segundo y más grande grupo estuvo conformado por los siete jugos restantes de Argentina y el total de muestras de

Brasil. Los resultados alcanzados estuvieron de acuerdo con los obtenidos mediante PCA, en el que se pudo diferenciar dos grandes grupos, con las mismas muestras de AR y BR agrupadas.

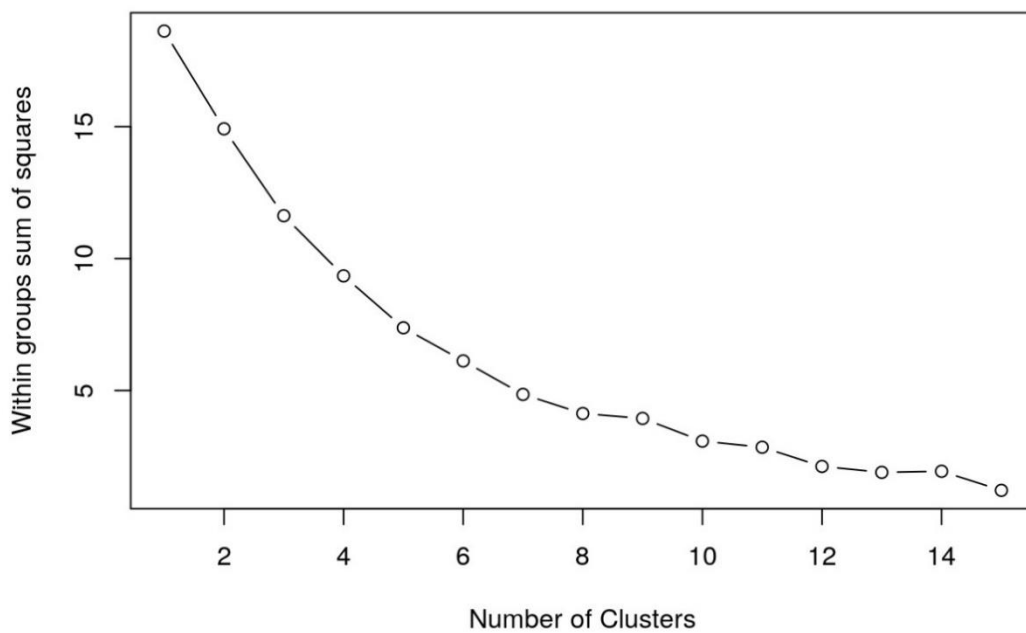
Además de aplicar HCA, se probó con el método de clúster o conglomerados por K-medias (en inglés, k-means) para dividir el conjunto de datos dado en un conjunto de k grupos (es decir, k clústeres) donde k representa el número de grupos establecidos por el analista como un requisito previo para aplicar el algoritmo. En la Fig. 4.13b se presenta el gráfico que muestra cómo la suma de los errores al cuadrado dentro del grupo (en inglés within-cluster sum of squared error WSS) disminuye a medida que se agrega un grupo (aumenta k) en k-medias. Este procedimiento se llevó a cabo para comprender cómo disminuye el error total de WSS a medida que aumentan los grupos y poder así establecer el valor de k óptimo para llevar a cabo el algoritmo de k-means. En el gráfico, no se observó un marcado quiebre o codo, por lo que en este trabajo, se aplicó la agrupación de k-medias considerando $k = 2$, ya que solo hay dos categorías de interés AR y BR. Para graficar las observaciones agrupadas en los 2 grupos, se presenta la Fig. 4.13c con un gráfico bidimensional de los clusters llamado "clusplot". Este gráfico muestra la componente 1 en el eje x, la componente 2 en el eje y, y los clústers se representan como elipses que contienen los objetos (muestras). Las elipses se construyen teniendo en cuenta el promedio y la matriz de covarianza de cada grupo, y su tamaño es tal que contienen todos los puntos de su grupo. Para conjuntos de datos de dimensiones superiores, se aplica una técnica de reducción de dimensiones antes de construir el gráfico. Después de analizar esta gráfica, se puede observar que en el grupo 2 la mayoría de las observaciones fueron muestras de jugo de uva de Argentina, excepto una muestra (punto 17) correspondiente a un jugo de Brasil. El grupo 1 estuvo formado por nueve muestras de Brasil y cinco de Argentina. Estos resultados muestran un mejor desempeño del algoritmo con respecto al análisis HCA.

El análisis exploratorio basado en la concentración de los elementos determinados, permitió una separación de las muestras de jugos de uva según su país de origen, con sólo unas pocas muestras que no se diferenciaron claramente.

a)



b)



c)

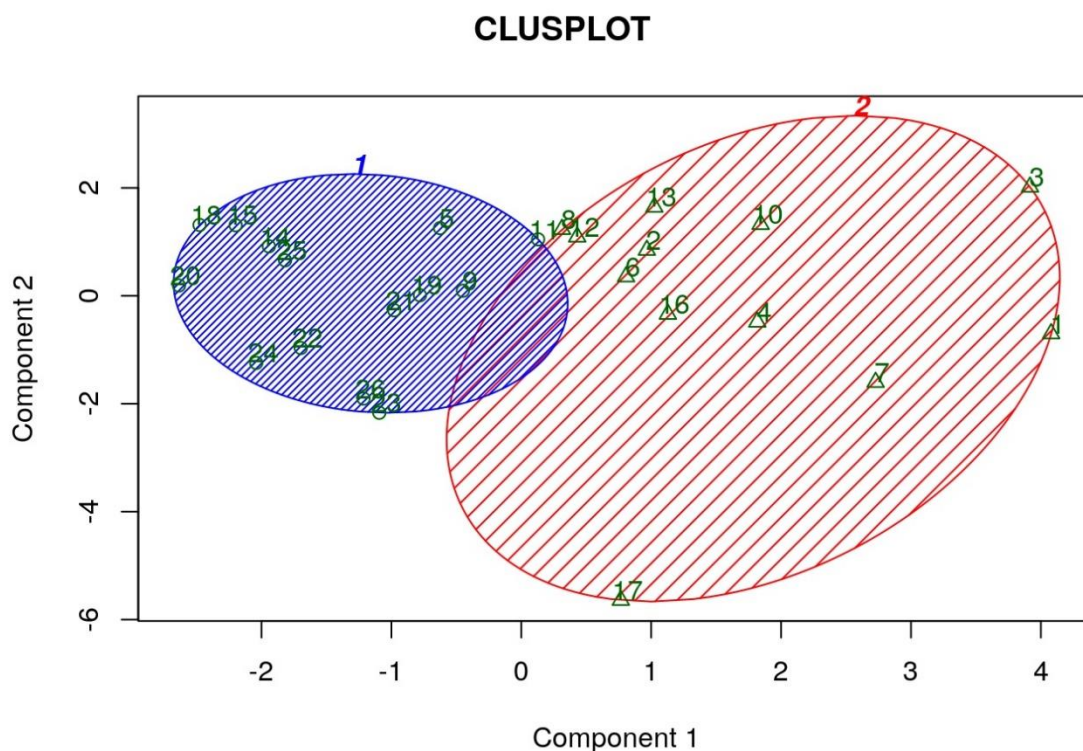


Figura 4.13. Dendrograma del método de clúster jerárquico (a). Gráfico de suma de los errores al cuadrado dentro del grupo (WSS) versus número de clústers (b). Clusplot para el análisis de k-medias (c).

4.2.1.4. Análisis clasificatorio

El análisis exploratorio mediante PCA y CA proporcionó buenos resultados en la diferenciación geográfica de los jugos de uva analizados en este trabajo, sin embargo, estas técnicas no pueden considerarse propiamente como métodos de clasificación, ya que se trata de herramientas estadísticas multivariadas no supervisadas. Por lo tanto, se emplearon los métodos supervisados como LDA y NNET para evaluar la posible clasificación geográfica de los jugos de uva de Argentina y Brasil, en función de los diversos elementos analizados, obteniendo un valor de accuracy de la performance de la clasificación.

Debido a las pocas muestras disponibles por país de origen para realizar el entrenamiento del modelo, no es recomendable utilizar un enfoque de división de los datos, ya que dividir los mismos en un conjunto de entrenamiento y prueba y simplemente validar,

no proporcionaría resultados representativos muy buenos. Por lo tanto, para este trabajo fue necesario utilizar otros enfoques, como bootstrapping (BTS) y validación cruzada (CV), especialmente adecuados para su uso en situaciones en las que hay disponible poca cantidad de muestras, como técnicas de validación para reducir la variación y el sobreajuste de los modelos. CV y BTS son métodos diseñados para estimar el rendimiento del clasificador utilizando las muestras disponibles mediante particionamiento (CV) o remuestreo con reemplazo (BTS) [24].

El método de bootstrapping crea muestras de bootstrap (subconjuntos aleatorios de datos de entrenamiento y de validación) al recolectar iterativamente observaciones del conjunto de datos inicial con reemplazo, es decir cientos o miles de repeticiones que implican que la misma observación puede participar más de una vez. Esto da como resultado una muestra con el mismo tamaño que el conjunto de datos original. Al usar BTS, es necesario elegir el número de repeticiones, en este trabajo se aplicó BTS con 100 repeticiones.

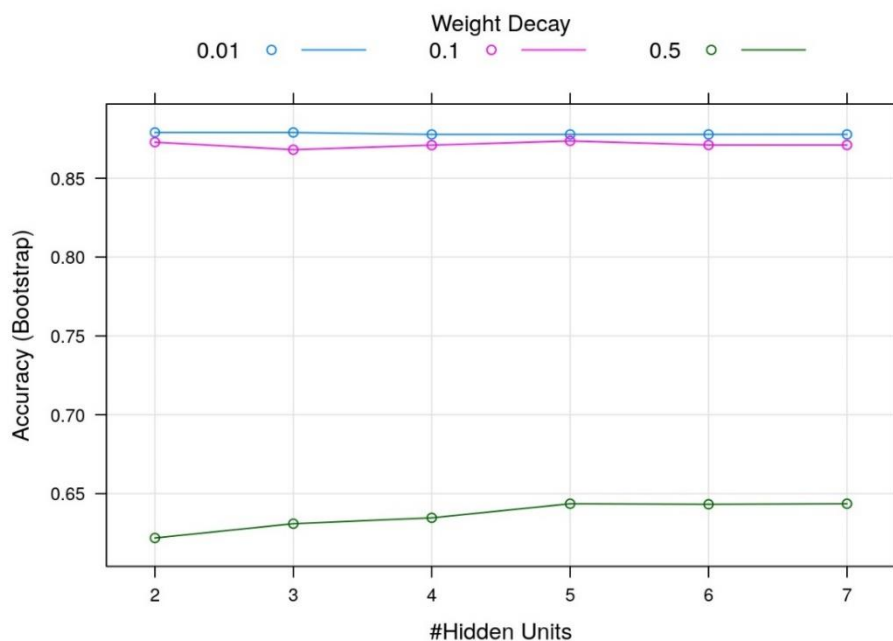
La CV implica dividir un conjunto de datos en subconjuntos, realizar el análisis en un subconjunto (llamado conjunto de entrenamiento) y validar el análisis en el otro subconjunto (llamado conjunto de validación). En k-fold CV, el conjunto de datos se divide aleatoriamente en k subconjuntos mutuamente excluyentes de aproximadamente el mismo tamaño. El clasificador es entrenado y probado k veces, y para estimar la precisión de los modelos, el número total de clasificaciones correctas se divide por el número de instancias en el conjunto de datos. En este trabajo k-fold CV se llevó a cabo para un $k = 4$ para los modelos.

En primer lugar, se aplicó el método clásico de análisis discriminante lineal, LDA, considerando todas las muestras de jugo de uva y se analizó su precisión. LDA con bootstrapping proporcionó un porcentaje de clasificación general del 75% y un valor kappa de 0,47. Un valor bajo de kappa no es un indicador confiable para este modelo. En el caso de LDA con 4-fold CV, los resultados de exactitud para la clasificación fueron del 81% y con un valor

kappa de 0,62. Se puede observar que no hubo una marcada diferencia entre los dos enfoques de validación para el modelo LDA.

Posteriormente se evaluó el método NNET para la clasificación geográfica de los jugos de uva. Aplicando BTS, inicialmente se entrenó el algoritmo considerando dos parámetros, el tamaño de las unidades ocultas o neuronas y el valor de la pérdida de peso (weight decay) que evitan que el modelo se vuelva demasiado complejo. Para la optimización, se usó el mayor valor de accuracy para seleccionar los valores óptimos para los dos parámetros del modelo. Los valores finales utilizados para el modelo fueron 2 neuronas y pérdida de peso de 0,01. La exactitud de la tasa general para este modelo fue del 87,3% con un valor kappa de 0,72. Cuando se realizó NNET con 4-fold CV los parámetros utilizados para el algoritmo fueron 2 neuronas y pérdida de peso de 0,1, siendo el porcentaje de la clasificación de 89,3% y el valor de kappa de 0,79. La Fig. 4.14 presenta dos gráficos con el rendimiento del método NNET cuando se variaron los parámetros de pérdida de peso y tamaño de las unidades ocultas para los dos métodos de validación. Finalmente, NNET con k-fold CV mostró el mejor resultado para el problema de clasificación geográfica de los jugos de uva.

a)



b)

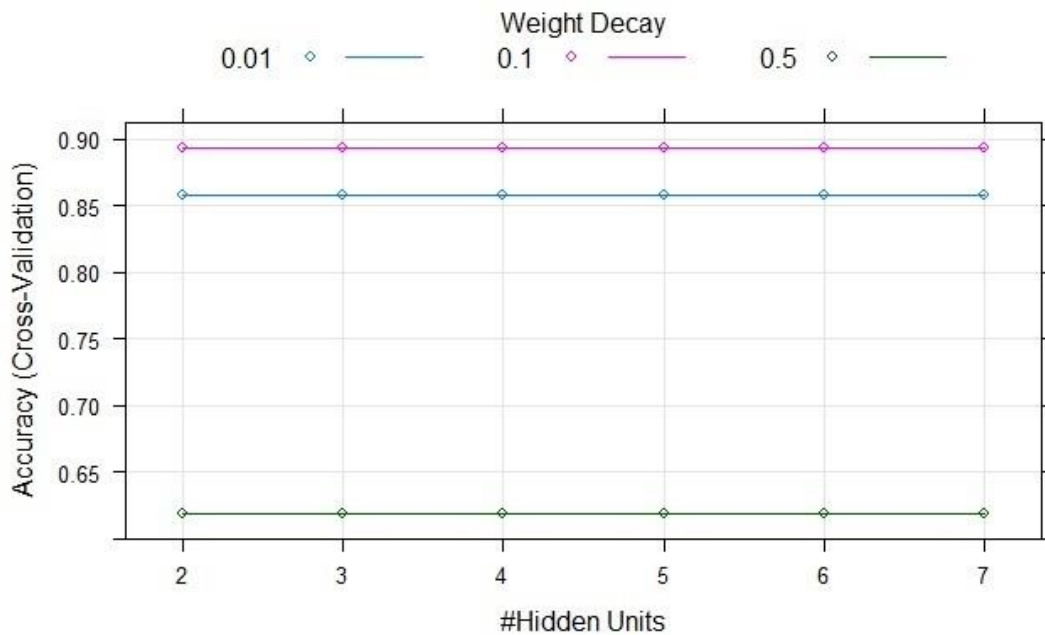
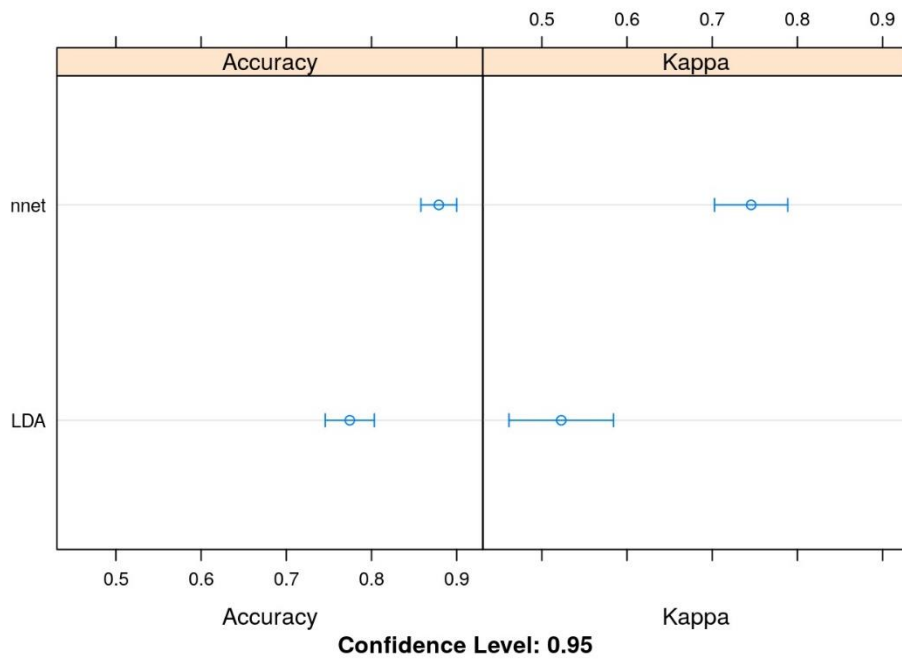


Figura 4.14. Rendimiento del modelo NNET al variar el tamaño de las unidades ocultas o neuronas y disminución de peso para los métodos de validación BTS (a) y CV (b).

Finalmente, para comparar el rendimiento de los dos métodos de clasificación, se presenta la Fig. 4.15, en la que se puede observar la distribución del valor de accuracy y de kappa para cada método de validación BTS (a) y CV (b). Por lo tanto, los límites alrededor de la media representan la amplitud de la dispersión de los resultados obtenidos para cada ciclo de remuestreo. En la figura se puede apreciar que los valores de accuracy y kappa resultaron mejor para el modelo NNET, indicando que es un método adecuado y versátil para la clasificación de este tipo de datos de concentraciones de jugos de uva, además se observa una gran dispersión del rendimiento de los modelos, principalmente de LDA con la aplicación de 4-fold CV.

a)



b)

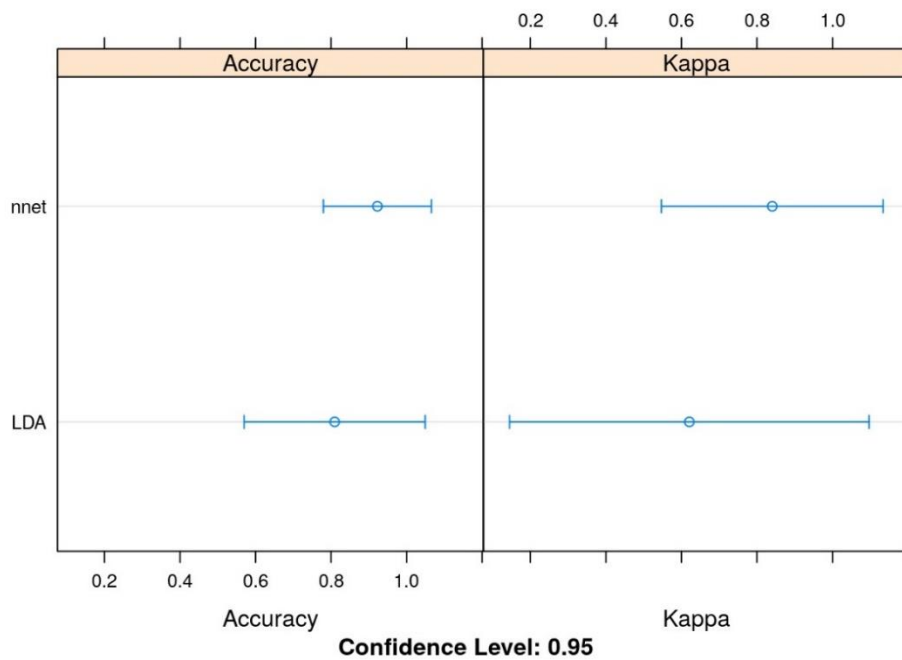


Figura 4.15. Gráfico de las distribuciones de “accuracy” y “kappa” para los métodos de validación BTS (a) y CV (b).

4.2.1.5. Conclusiones

En este trabajo se describe la aplicación del análisis ICP-MS combinado con técnicas quimiométricas para la clasificación geográfica de jugos comerciales obtenidos de Argentina y Brasil. Los niveles de concentración de once elementos se interpretaron utilizando métodos no supervisados como PCA y CA, que proporcionaron un enfoque exploratorio y simple para evaluar el origen y, por lo tanto, la autenticidad de las muestras de jugo de uva. Los resultados de estas técnicas permitieron una separación de las muestras de jugos de uva según su país de origen. Adicionalmente, se evaluaron dos algoritmos de clasificación, LDA y NNET, mediante dos métodos de validación, de particionamiento (CV) o remuestreo con reemplazo (BTS). Estos métodos proporcionan una buena capacidad de predicción de origen para las muestras, resultando el mejor algoritmo para la clasificación NNET con 4-fold CV, con un valor de accuracy de casi el 90%.

En conclusión, utilizando diferentes técnicas quimiométricas fue posible distinguir entre muestras de jugo de uva de Argentina y Brasil teniendo en cuenta su contenido elemental. Las concentraciones de elementos medidos en los jugos de uva resultaron ser marcadores relevantes de la procedencia de las muestras, teniendo un excelente potencial para discriminar el origen de los jugos de uva comerciales asegurando la calidad de los productos.

4.2.1.6. Referencias

- [1] Codex Alimentarius Commission, Codex General Standard for Fruit Juices and Nectars, Codex Stand 247. (2005) 1–19. http://www.fao.org/fao-who-codexalimentarius/sh-proxy/en/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252Fstandards%252FCXS%2B247-2005%252FCXS_247e.pdf (accessed August 28, 2019).
- [2] R.E.S. Froes, W.B. Neto, N.O.C. e. Silva, R.L.P. Naveira, C.C. Nascentes, J.B.B. da Silva, Multivariate optimization by exploratory analysis applied to the determination of

- microelements in fruit juice by inductively coupled plasma optical emission spectrometry, *Spectrochim. Acta Part B At. Spectrosc.* 64 (2009) 619–622. doi:10.1016/j.sab.2009.06.006.
- [3] R.L. Bhardwaj, U. Nandal, A. Pal, S. Jain, Bioactive compounds and medicinal properties of fruit juices, *Fruits.* 69 (2014) 391–412. doi:10.1051/fruits/2014027.
- [4] A. Szymczycha-Madeja, M. Welna, D. Jedryczko, P. Pohl, Developments and strategies in the spectrochemical elemental analysis of fruit juices, *TrAC Trends Anal. Chem.* 55 (2014) 68–80. doi:10.1016/j.trac.2013.12.005.
- [5] I.M. Toaldo, O. Fogolari, G.C. Pimentel, J.S. de Gois, D.L.G. Borges, V. Caliari, M. Bordignon-Luiz, Effect of grape seeds on the polyphenol bioactive content and elemental composition by ICP-MS of grape juices from *Vitis labrusca* L., *LWT - Food Sci. Technol.* 53 (2013) 1–8. doi:10.1016/j.lwt.2013.02.028.
- [6] Cámara Argentina de Fabricantes y Exportadores de Mosto, El 90% del jugo de uva argentino está destinado a la exportación, (2019). <http://www.prensa.mendoza.gov.ar/el-90-del-jugo-de-uva-argentino-esta-destinado-a-la-exportacion/> (accessed July 15, 2019).
- [7] M. Díaz-García, M.R. Castellar, J.M. Obón, Juices and Non-Alcoholic Beverages, in: *Compr. Anal. Chem.*, 2013: pp. 439–459. doi:10.1016/B978-0-444-59562-1.00017-7.
- [8] D. Granato, M. de Magalhães Carrapeiro, V. Fogliano, S.M. van Ruth, Effects of geographical origin, varietal and farming system on the chemical composition and functional properties of purple grape juices: A review, *Trends Food Sci. Technol.* 52 (2016) 31–48. doi:10.1016/j.tifs.2016.03.013.
- [9] J. Guo, T. Yue, Y. Yuan, Y. Wang, Chemometric Classification of Apple Juices According to Variety and, *J. Agric. Food Chem.* 61 (2013) 6949–6963.
- [10] C. Jiménez-Jorquera, M. Gutiérrez-Capitán, Electronic Tongues Applied to Grape and Fruit Juice Analysis, in: *Electron. Noses Tongues Food Sci.*, Elsevier, 2016: pp. 189–198.

- doi:10.1016/B978-0-12-800243-8.00019-6.
- [11] A. Dehelean, D.A. Magdas, Analysis of Mineral and Heavy Metal Content of Some Commercial Fruit Juices by Inductively Coupled Plasma Mass Spectrometry, *Sci. World J.* 2013 (2013) 1–6. doi:10.1155/2013/215423.
- [12] P. Pohl, D. Jedryczko, A. Dzimitrowicz, A. Szymczycha-Madeja, M. Welna, P. Jamroz, Determination of Elements in Fruit Juices, in: *Fruit Juices*, Elsevier, 2018: pp. 739–761. doi:10.1016/B978-0-12-802230-6.00035-7.
- [13] M. Likar, K. Vogel-Mikuš, M. Potisek, K. Hančević, T. Radić, M. Nečemer, M. Regvar, Importance of soil and vineyard management in the determination of grapevine mineral composition, *Sci. Total Environ.* 505 (2015) 724–731. doi:10.1016/j.scitotenv.2014.10.057.
- [14] A. Szymczycha-Madeja, M. Welna, Evaluation of a simple and fast method for the multi-elemental analysis in commercial fruit juice samples using atomic emission spectrometry, *Food Chem.* 141 (2013) 3466–3472. doi:10.1016/j.foodchem.2013.06.067.
- [15] E.M. Borges, D.A. Volmer, E. Brandelero, J.M.L.N. Gelinski, M. Gallimberti, F. Barbosa, Monitoring the Authenticity of Organic Grape Juice via Chemometric Analysis of Elemental Data, *Food Anal. Methods.* 9 (2016) 362–369. doi:10.1007/s12161-015-0191-x.
- [16] V. Uríčková, J. Sádecká, Determination of geographical origin of alcoholic beverages using ultraviolet, visible and infrared spectroscopy: A review, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 148 (2015) 131–137. doi:10.1016/j.saa.2015.03.111.
- [17] L. Tormen, D.P. Torres, I.M. Dittert, R.G.O. Araújo, V.L.A. Frescura, A.J. Curtius, Rapid assessment of metal contamination in commercial fruit juices by inductively coupled mass spectrometry after a simple dilution, *J. Food Compos. Anal.* 24 (2011) 95–102. doi:10.1016/j.jfca.2010.06.004.

- [18] C. Dani, L.S. Oliboni, D. Prá, D. Bonatto, C.E.I. Santos, M.L. Yoneama, J.F. Dias, M. Salvador, J.A.P. Henriques, Mineral content is related to antioxidant and antimutagenic properties of grape juice, *Genet. Mol. Res.* 11 (2012) 3154–3163. doi:10.4238/2012.September.3.4.
- [19] M.P. Fabani, M.E. Toro, F. Vázquez, M.P. Díaz, D.A. Wunderlin, Differential absorption of metals from soil to diverse vine varieties from the valley of tulum (Argentina): Consequences to evaluate wine provenance, *J. Agric. Food Chem.* 57 (2009) 7409–7416. doi:10.1021/jf901572k.
- [20] A. Miele, L.A. Rizzon, S.C. do N. de Queiroz, C. Gianello, Physicochemical composition, minerals, and pesticide residues in organic grape juices, *Food Sci. Technol.* 35 (2015) 120–126. doi:10.1590/1678-457X.6540.
- [21] L.A. Rizzon, M. Link, Composição do suco de uva caseiro de diferentes cultivares, *Ciência Rural.* 36 (2006) 689–692. doi:10.1590/S0103-84782006000200055.
- [22] L.A. Rizzon, A. Miele, Analytical characteristics and discrimination of Brazilian commercial grape juice, nectar, and beverage, *Food Sci. Technol.* 32 (2012) 93–97. doi:10.1590/S0101-20612012005000015.
- [23] I.M. Toaldo, F.A. Cruz, T.D.L. Alves, J.S. de Gois, D.L.G. Borges, H.P. Cunha, E.L. da Silva, M.T. Bordignon-Luiz, Bioactive potential of *Vitis labrusca* L. grape juices from the Southern Region of Brazil: Phenolic and elemental composition and effect on lipid peroxidation in healthy subjects, *Food Chem.* 173 (2015) 527–535. doi:10.1016/j.foodchem.2014.09.171.
- [24] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proc. 14th Int. Jt. Conf. Artif. Intell.*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995: pp. 1137–1143. doi:10.1067/mod.2000.109032.

4.3. CLASIFICACIÓN GEOGRÁFICA DE VINOS

4.3.1. Autenticación geográfica de vinos tintos de Argentina mediante análisis multielemental y quimiométrico clasificatorio

4.3.1.1. Introducción

El vino es una bebida alcohólica resultante exclusivamente de la fermentación alcohólica parcial o completa de uvas frescas o mosto de uva [1]. Esta bebida tiene una gran relevancia, ya que su producción es una importante actividad en la economía mundial [2,3]. Aunque la producción de vino se ha mantenido con altos y bajos en los últimos años, la tendencia del consumo a nivel mundial, ha aumentado ligeramente. Según la Organización Internacional de la Viña y el Vino (OIV), la producción y el consumo de vino se estimaron en alrededor de 260 y 244 millones de hectolitros respectivamente para el año 2019, lo que destaca la importancia de esta bebida en el comercio mundial [4]. Dentro de la producción mundial, se destaca la de Argentina, ocupando el quinto puesto según su nivel de producción en el año 2019.

Desde un punto de vista químico, el vino es una matriz realmente compleja compuesta de moléculas de diferente naturaleza y estructura (compuestos volátiles como el etanol, proteínas, aminoácidos, carbohidratos, compuestos fenólicos e inorgánicos) [5]. Esta compleja mezcla está influenciada por varios factores como el clima, la ubicación geográfica, la selección de variedades de uva y el momento de la cosecha, el proceso de fermentación, las condiciones de almacenamiento y envejecimiento, así como las prácticas vitivinícolas generales [6,7]. Estas características hacen del vino una bebida única y valiosa con un valor agregado distintivo para el consumidor y el productor, basado en su origen geográfico y, por lo tanto, en su "terroir" asociado [5].

En el mundo de la viticultura, el término francés "terroir" (de la palabra terre, que significa suelo) se define como la interacción entre el entorno físico y biológico con las vides, en el que la geología, geomorfología, suelo, condiciones climáticas, biología de la vid y la

actividad humana proporciona características distintivas y de alta calidad para los productos originados en un área específica [8–10]. En este marco, los sistemas de clasificación geográfica surgen como indicadores de autenticidad y calidad del vino y se establecen en la mayoría de los países productores mundiales de vino. Además, tiene una gran importancia en el área económica ya que el valor comercial de un vino se basa en esos parámetros intensamente relacionados con la historia y el origen del producto.

La verificación y evaluación de la autenticidad y calidad del vino requieren evaluar tres factores claves como la declaración de origen, la composición real de acuerdo con la etiqueta (variedad de uva y producción enológica) y la cosecha [11]. Hoy en día existe un conjunto de procedimientos y marcos legales estrictos para la trazabilidad del vino y el control de calidad con el fin de garantizar la seguridad y excelencia de los vinos estrictamente vinculada a un área geográfica particular y a métodos específicos de producción; estos son las designaciones: Especialidad Tradicional Garantizada (TSG), Denominación de Origen Protegida o Controlada (PDO) e Indicación Geográfica Protegida (PGI) [12].

Los vinos con denominación de origen deben producirse, procesarse y prepararse en una región específica utilizando métodos tradicionales que les confieran un alto valor económico. En este sentido, la existencia de diversas técnicas analíticas ha sido útil para clasificar las muestras de vino de acuerdo con su terruño, origen geográfico, variedad y edad. Asimismo, los métodos quimiométricos tienen como objetivo identificar marcadores únicos o grupos de marcadores que caracterizan la autenticidad.

Las técnicas quimiométricas se consideran herramientas útiles para la evaluación de la autenticidad y la trazabilidad y para la detección de cualquier adulteración de alimentos. Estos métodos se aplican sobre grandes cantidades de datos de características multivariadas, es decir, el análisis de más de una variable de muestra, con el objetivo de detectar un patrón o huella dactilar que pueda usarse para inferir la naturaleza auténtica de un producto

alimenticio o bebida y clasificar productos según origen, calidad, variedad, tipo u otras características [13,14].

En el contexto del vino, las herramientas quimiométricas se están volviendo cada vez más populares y significativas para la evaluación de la autenticidad y la calidad de esta bebida. La combinación de análisis instrumentales con estadísticas multivariadas ha permitido la clasificación exitosa de los vinos de acuerdo con su origen geográfico, variedades y ciertas características del proceso de elaboración del vino a nivel mundial.

El patrón químico inorgánico del vino es una excelente huella geoquímica que se usa cada vez más para evaluar el origen geográfico de los vinos. Esto se debe a que la composición mineral puede considerarse como una consecuencia de las características de la geoquímica local, el clima y las prácticas culturales [15–17]. Además, dado que no todos los elementos se metabolizan o modifican durante el proceso de elaboración del vino, estos se pueden considerar como buenos marcadores. Algunos minerales en el suelo se absorben a través de las raíces de la vid y pasan a las uvas, y posteriormente se transfieren al vino durante el proceso de vinificación. Las técnicas de vinificación no cambian en gran medida la composición mineral del vino derivado de las uvas, lo que indica que el factor principal que afecta la composición mineral del vino es el origen geográfico del vino [18]. Las variables ideales para la clasificación del vino son los elementos completamente no afectados por el proceso de elaboración del vino, que por lo tanto reflejan la peculiaridad geoquímica del suelo. No es fácil individualizar tales variables, por lo que en muchos casos, el perfil multielemental para la clasificación geográfica se ha utilizado como una especie de enfoque no dirigido [18–20].

Por consiguiente, el objetivo de este trabajo fue evaluar el uso potencial de varias herramientas quimiométricas combinadas con un análisis multielemental realizado mediante la técnica de ICP-MS para la diferenciación según su origen geográfico de diferentes muestras de vino tinto procedentes de las tres regiones vitivinícolas de Argentina más importantes, Mendoza, Noroeste argentino y la Patagonia.

4.3.1.2. Concentración multielemental de vinos

En la Tabla 4.7 se muestra la concentración de los diferentes elementos determinados mediante ICP-MS en las muestras de vino tinto procedente de tres grandes regiones productoras de vino, Mendoza, Patagonia y Noroeste Argentino, expresadas como valores medios con su correspondiente desviación estándar en $\mu\text{g/L}$.

Tabla 4.7. Concentración elemental de las muestras de vino según su región de procedencia.

Elemento	LOD ($\mu\text{g/L}$)	Mendoza (n=96)		Patagonia (n=25)		Noroeste (n=13)	
		Media	DE	Media	DE	Media	DE
Ag	0,05	2,4	16,3	0,54	1,04	0,30	0,25
Al	10	259	81	592	210	386	126
As	0,05	6,5	4,0	7,6	1,9	10,3	3,2
Ba	0,05	38,1	17,4	54,5	22,0	82,9	24,3
Be	0,05	0,31	0,24	0,36	0,17	0,32	0,28
Cd	0,05	0,21	0,16	0,36	0,46	0,26	0,18
Co	0,05	1,4	0,65	2,1	0,84	1,7	0,62
Cr	0,05	12,8	2,9	16,3	4,9	17,6	6,3
Cs	0,05	1,5	1,3	2,6	2,8	3,2	2,3
Cu	0,5	107	57,1	145	48,1	112	37,4
Fe	5	1545	471	1731	672	1884	708
Ga	0,05	0,09	0,07	0,23	0,26	0,13	0,06
Hg	0,05	1,4	0,94	0,76	0,32	1,3	0,69
Li	0,25	109	41,4	58,1	22,4	167	92,4
Mn	0,25	818	206	779	272	1234	348
Mo	0,05	4,0	3,0	4,0	1,9	6,0	3,9
Ni	0,05	17,2	7,6	19,2	20,9	15,9	5,6
Pb	0,05	10,1	8,4	8,9	7,1	14,1	9,6
Pd	0,05	2,6	8,2	2,1	1,6	1,5	0,68
Rb	0,05	927	376	1009	313	1295	733
Sb	0,05	10,2	10,3	7,0	4,2	7,4	8,2
Se	0,05	2,8	1,6	1,8	0,45	3,2	1,4
Sn	0,05	17,8	20,3	17,8	15,5	10,8	4,8
Sr	0,05	1099	293	961	186	822	242
Tl	0,05	0,24	0,46	0,26	0,20	0,15	0,09
V	0,05	6,7	25,9	1,3	1,5	0,49	0,23
Zn	0,1	415	198	551	233	506	254

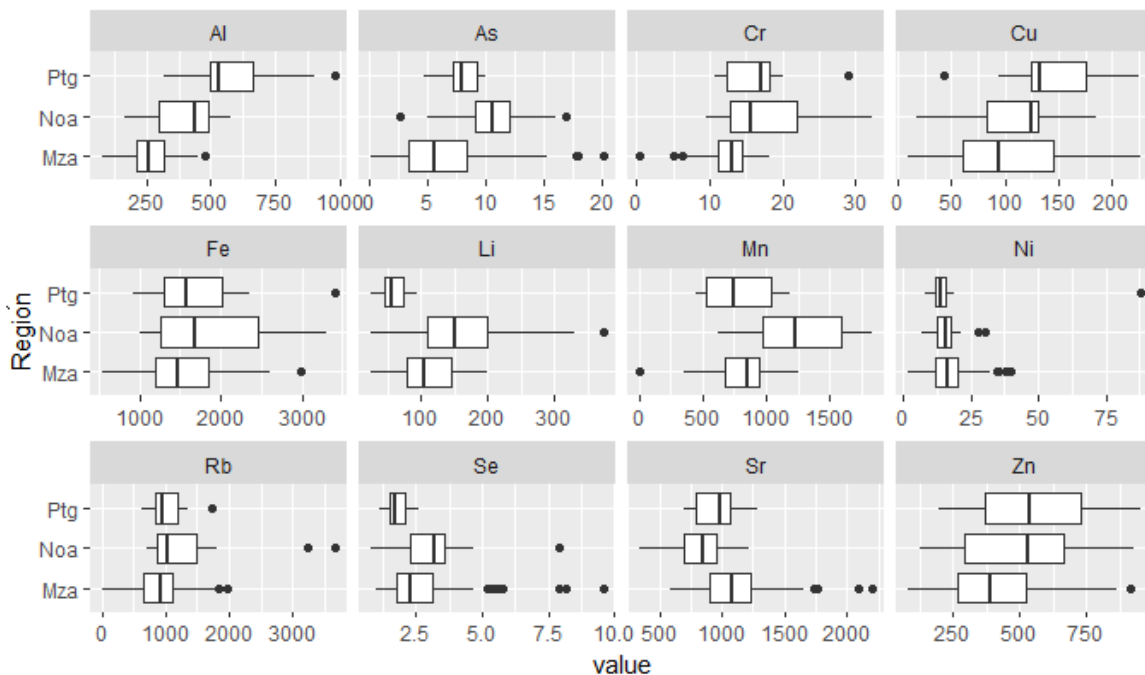
A partir del análisis de los resultados mostrados en la Tabla 4.7, se observa que Fe es el elemento más abundante encontrado en los vinos tintos, seguido por Rb, Mn, y Sr. El contenido medio de Fe más alto se obtuvo en los vinos procedentes de la región del Noroeste Argentino (1884 $\mu\text{g/L}$) mientras que la concentración media más baja de este elemento se encontró en los vinos de Mendoza (1545 $\mu\text{g/L}$). La concentración de Rb hallada estuvo en el rango de 927 $\mu\text{g/L}$ y 1295 $\mu\text{g/L}$, mientras que las de Mn y Sr estuvieron en el rango de 779 - 1234 $\mu\text{g/L}$ y 822 - 1099 $\mu\text{g/L}$, respectivamente. Estas concentraciones estuvieron en concordancia con lo reportado por otros autores en muestras de vinos tintos y en general en el mismo orden de magnitud [19,21–23]. Con respecto a otros trabajos realizados sobre muestras de vinos tintos, se encontraron diferencias en los niveles de estos elementos [18,24–26].

Los siguientes elementos encontrados en niveles traza, en el rango de concentraciones de 1000 $\mu\text{g/L}$ a 100 $\mu\text{g/L}$, fueron Al, Zn, Li, y Cu, mientras que Ba, Ni, Sn, Cr, Pb, As, Sb, V, Mo, Se, Pd, Cs, Ag, y Co se hallaron en concentraciones decrecientes por debajo del rango de los 100 microgramos y en concentraciones mayores al microgramo. Finalmente los elementos Be, Cd, Ga, Hg y Tl fueron hallados en concentraciones menores del microgramo por litro.

La concentración hallada de los elementos está en concordancia con lo expuesto por Pohl, quién detalla que los principales elementos del vino, presentes en niveles entre 10 y 1000 mg/L son Ca, K, Na y Mg, seguidos de elementos menores, tales como Al, Fe, Cu, Mn, Rb, Sr y Zn, presentes en el rango de 0.1-10 mg/L. Los elementos traza están presentes en el rango ppt (0.1–1000 $\mu\text{g/L}$), e incluyen Ba, Cd, Co, Cr, Li, Ni, Pb y V entre otros [27]. Algunos de estos elementos traza, como Fe, Zn, Cu, Cr y Se, son esenciales para el organismo humano, ya que forman parte integral de una o más enzimas involucradas en un proceso metabólico o bioquímico. En cantidades moderadas, el consumo de vino permite obtener estos elementos que son esenciales para el organismo humano, pero a su vez puede ser fuente de elementos tales como As, Cd, Cr, Hg y Pb, que deben controlarse debido a su potencial toxicidad [21,25].

La Organización Internacional de la Viña y el Vino (OIV) establece los límites máximos aceptables en el vino para los siguientes elementos: As (0,2 mg/L); B (80 mg/L); Cu (1 mg/L); Pb (0,15 mg/L); Cd (0,01 mg/L); Br (1 mg/L); Ag (0,1 mg/L) y Zn (5 mg/L). Todas las muestras de vino analizadas presentaron niveles de los elementos tóxicos como As, Cu, Pb, Cd, Ag y Zn por debajo de la concentración máxima permitida en vinos [28].

Con respecto al efecto en el vino, algunos elementos en ciertas concentraciones pueden causar defectos como turbidez, precipitación y oxidación, afectando las características organolépticas del vino, incluidos el sabor, la frescura, el aroma, el color y el sabor, principalmente durante la fermentación, maduración y almacenamiento del vino. Sin embargo, los metales minoritarios (Cu, Fe, Mn, Zn) y algunos metales traza también son favorables para las levaduras [27].



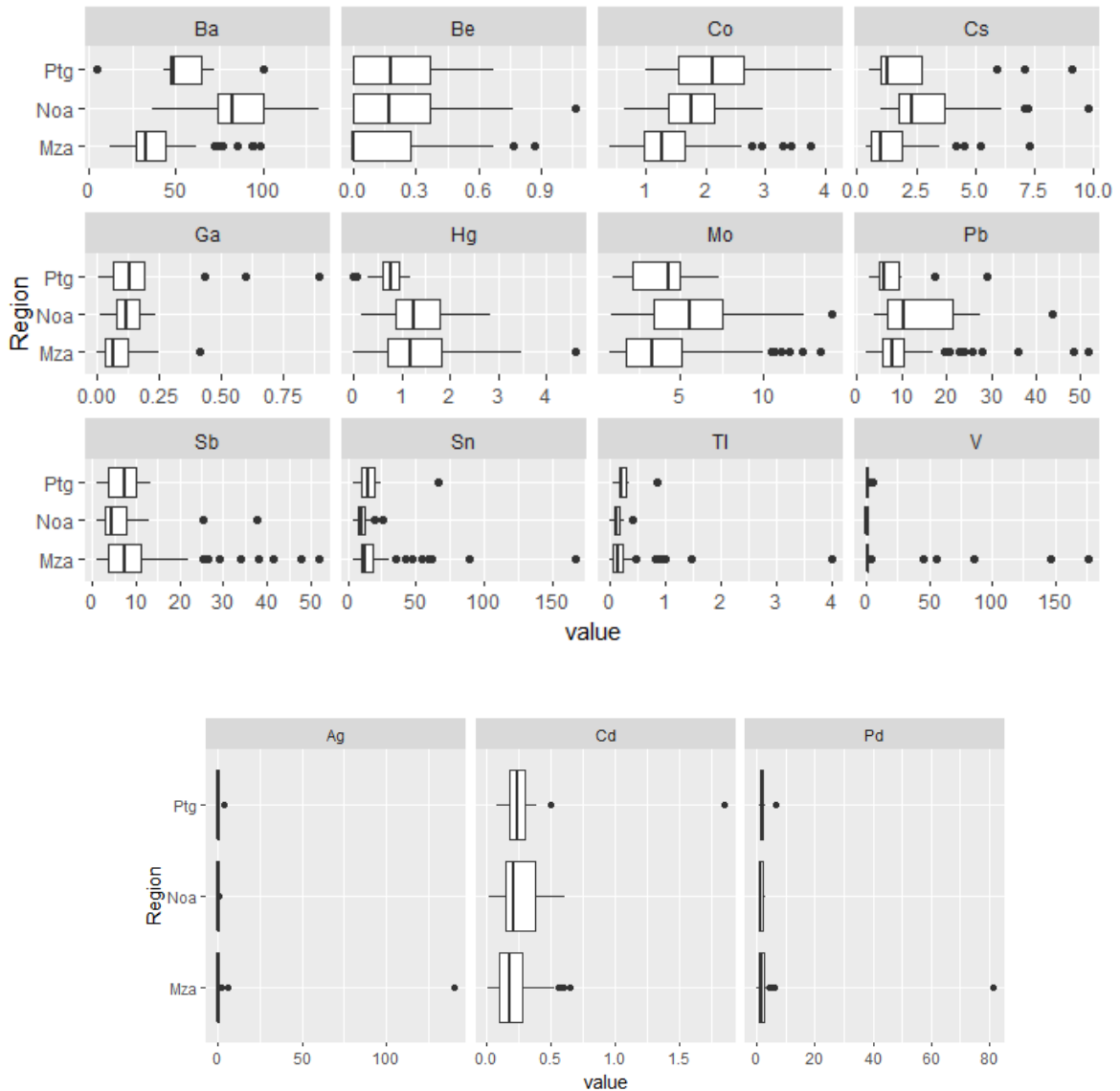


Figura 4.16. Gráficos Box-plot de los elementos determinados en vinos tintos según su procedencia (Ptg: Patagonia, Noa: Noroeste Argentino, Mza: Mendoza).

Para poder tener una mejor visualización de los datos que permita comparar las distribuciones de las concentraciones de los diferentes elementos determinados en las muestras de vino según su origen, se presenta en la Figura 4.16 los gráficos de caja y bigotes (box and whisker plot en inglés). La comparación de las concentraciones de los elementos analizados en las muestras de vino tinto teniendo en cuenta el origen geográfico, muestra que con respecto a los elementos encontrados en mayores concentraciones (Fe, Rb, Mn y Sr), los vinos presentaron concentraciones similares de los elementos según la región, presentando

una leve tendencia a mayores concentraciones y dispersión de los datos en los vinos de la región del Noroeste Argentino para los elementos Fe, Rb y Mn y una mayor concentración de Sr en los vinos de la provincia de Mendoza. Con respecto a los demás elementos determinados, se observa diferencia en la dispersión de las concentraciones de los elementos Al, As, Cu, Li, Ba y Co, según el lugar de procedencia de los vinos.

4.3.1.3. Análisis exploratorio

A partir de los datos de concentración de los diferentes elementos determinados en las muestras de vino tinto, se realizó una caracterización quimiométrica básica mediante el método de PCA para un análisis estadístico exploratorio. Como ya se ha mencionado anteriormente, este método permite la determinación de las principales fuentes de variabilidad en el conjunto de datos, y muestra la agrupación natural de los objetos estudiados en el plano o en un espacio en 3D de las componentes principales más importantes, que son creadas por una combinación lineal no correlacionada de todas las variables originales.

Para llevar a cabo el análisis de PCA, se tuvieron en cuenta los datos de concentración de los 27 elementos analizados en las muestras de vino. El contenido total de información dado por los componentes principales se expresa en forma de porcentaje acumulado de la varianza total. Cuando se utilizaron 27 variables, las primeras dos componentes principales (PC) representaron casi el 30% de la variabilidad explicada. La primera componente principal (PC1) representó el 19,6% de la varianza total y la siguiente componente principal (PC2) 8%, requiriéndose de 11 PCs para lograr el 71,1% y 14 PCs para llegar a casi el 80%, de resumen de la varianza, lo que indica que prácticamente no existen variaciones independientes entre las variables consideradas (Fig. 4.17).

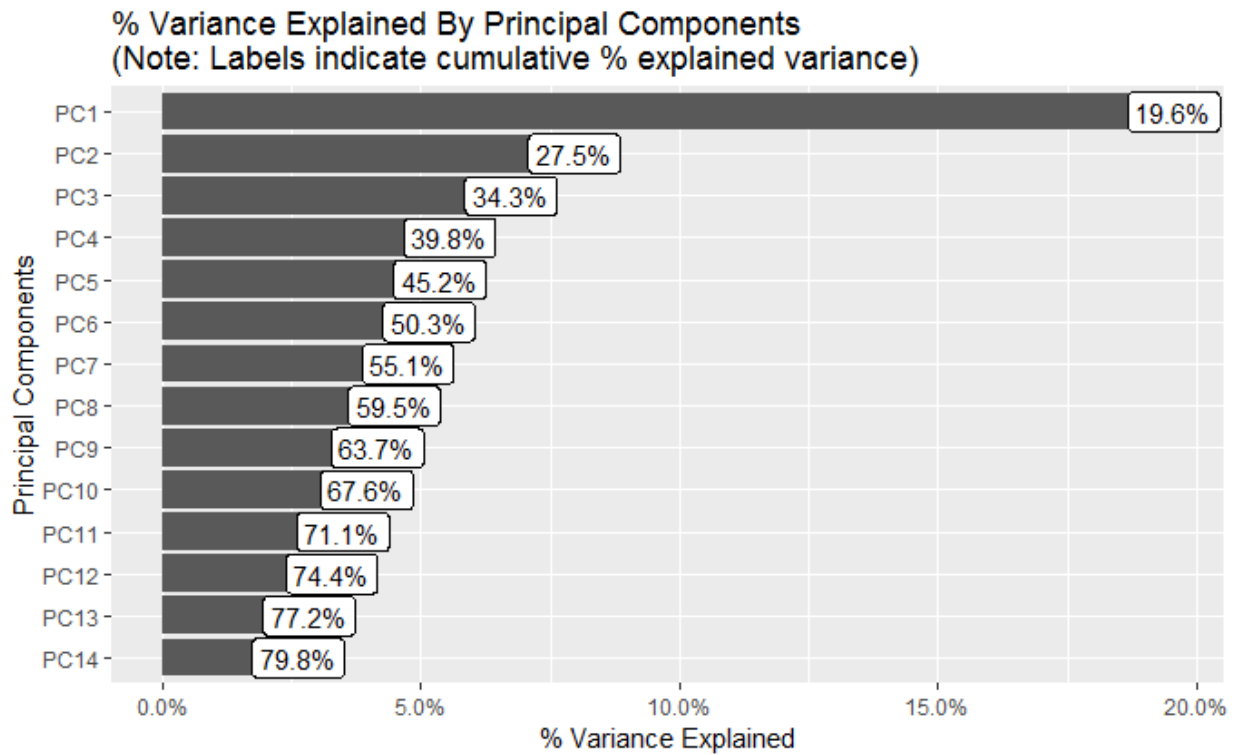
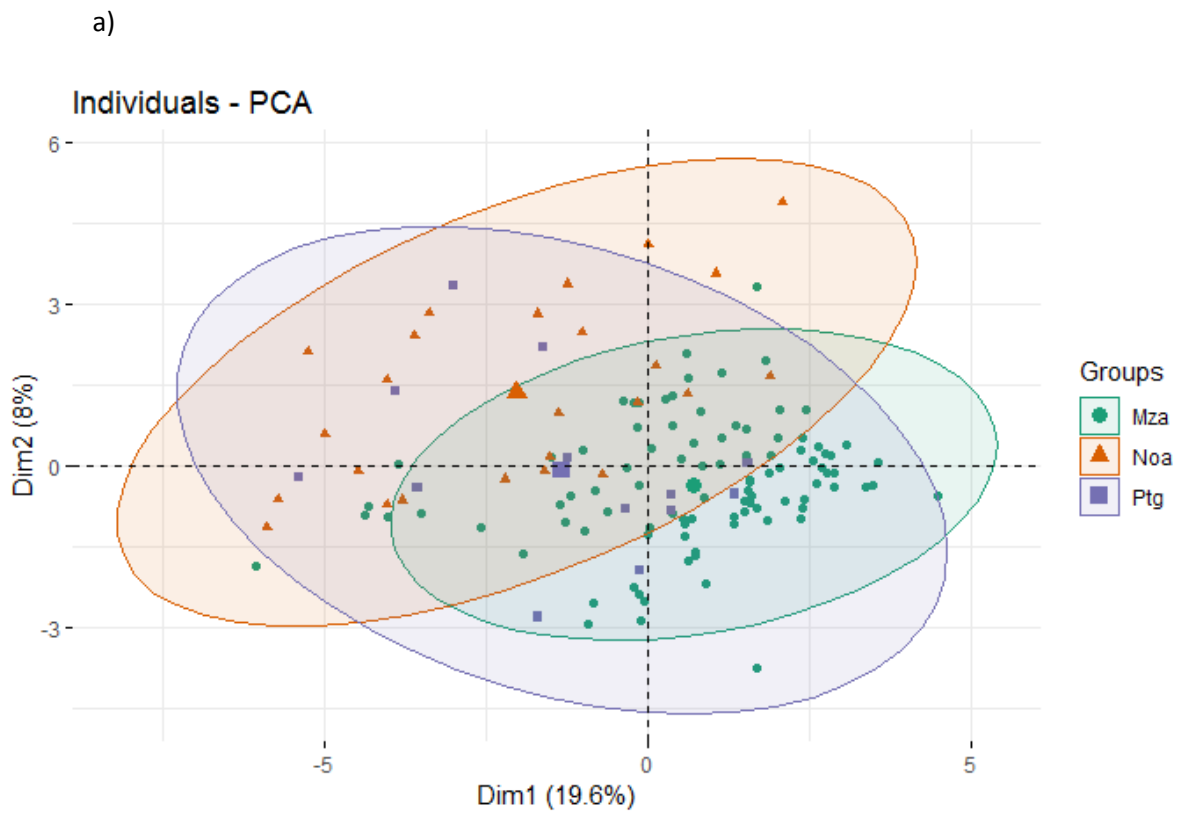


Figura 4.17. Varianza explicada por las componentes principales.

En la Figura 4.18 se muestran los dos gráficos más importantes de PCA, el de observaciones (scores) y variables (loadings) de las PC1 vs PC2, donde se muestra la agrupación natural de las 134 muestras de vino tinto de Argentina y la orientación de las variables respectivamente. La Fig. 4.18a muestra el gráfico de scores u observaciones de las dos primeras componentes principales, en esta, se puede observar una superposición entre las puntuaciones correspondientes a las diferentes muestras identificadas según su origen. Sin embargo, se puede observar un agrupamiento de las muestras de Mendoza (Mza) en la intersección de los ejes del gráfico, diferenciándose de las muestras de vino del Noroeste Argentino (Noa) que se encuentran más dispersas, mostrando una tendencia a scores negativos de los componentes. Las muestras de la región de la Patagonia (Ptg) aparecen completamente solapadas en las dos regiones anteriores.



b)

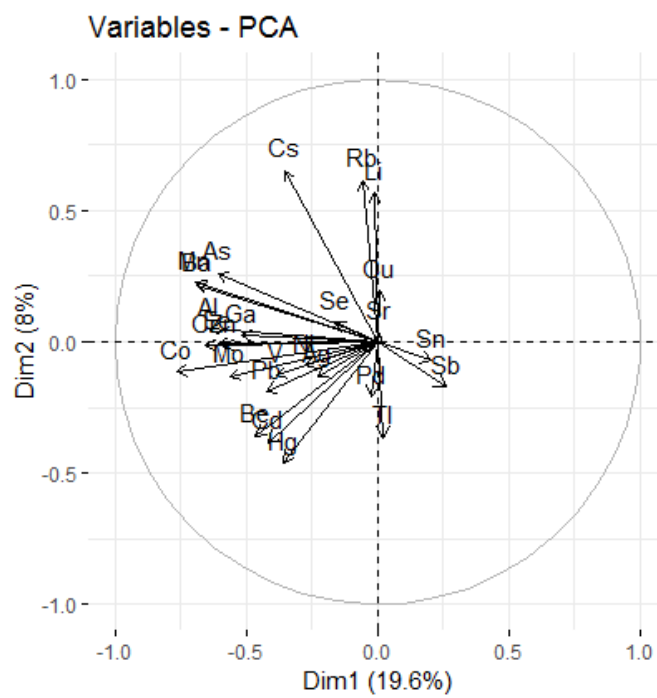


Figura 4.18. Gráfico PCA de a) observaciones y b) variables de la PC1 vs. PC2 de los vinos.

La Figura 4.18b presenta el gráfico de loadings en el plano PC1 vs.PC2, que ilustra la orientación de las variables (concentraciones de elementos) con respecto a las componentes principales más informativas PC1 y PC2. Se observa que la primera componente principal está fuertemente asociada con los valores de los elementos: Co, Ba, Mn, Cr, Al, As, Fe y Zn indicando concentraciones más altas en las muestras que mostraron puntuaciones negativas. Por otro lado, Cs, Rb y Li fueron las variables dominantes en la segunda componente principal, mientras que las muestras con puntuaciones positivas en PC2 corresponden a altas concentraciones de estos elementos. Igualmente las variables aparecen muy poco representadas en las primeras PCs, no brindando mucha diferenciación a la hora de la clasificación de las muestras de vino.

En resumen, los resultados obtenidos por PCA muestran que existe solapamiento de las muestras de vino tinto procedentes de las tres regiones vitivinícolas de Argentina por lo que no pudieron resolverse con este método quimiométrico no supervisado, necesitando de algoritmos quimiométricos más complejos para lograr la clasificación geográfica de las muestras de vino en estudio.

4.3.1.4. Análisis clasificatorio

Para poder proceder a la clasificación geográfica de las muestras de vino tinto, se emplearon técnicas de reconocimiento de patrones supervisadas, que implican principalmente la clasificación de objetos en grupos o categorías creando reglas de clasificación [29]. Se aplicaron cuatro métodos quimiométricos supervisados: el clásico LDA, k-NN, SVM y RF. Para ello, la matriz de datos de las concentraciones de los elementos, se dividió aleatoriamente en un conjunto de entrenamiento (70% de los objetos de toda la matriz de datos), con muestras de clase conocidas para calcular un clasificador, y el conjunto de prueba o testeo (30%) que contiene las muestras restantes no incluidas en el entrenamiento y que sirven para validar el rendimiento de los modelos.

Los métodos k-NN, SVM y RF, requirieron optimizar algunos parámetros para construir el modelo antes de su evaluación como herramienta de predicción geográfica de los vinos tintos teniendo en cuenta como criterio el mayor valor de accuracy para seleccionar el mejor modelo. Cada uno de los modelos fue entrenado usando una validación cruzada de k iteraciones o k -fold cross-validation (repetida n veces) en el conjunto de entrenamiento para construir los diferentes clasificadores. Este procedimiento se repitió n veces, por lo que cada subconjunto se utilizó para probar al menos una vez. En este trabajo para el método de remuestreo se empleó 5-fold-CV repetido 50 veces ya que al contar con pocas muestras en el grupo de entrenamiento y testeo de esta manera se logró la mejor performance de los modelos.

Para la optimización de k-NN, debido a su simplicidad, sólo se necesitó el ajuste de un parámetro, k , correspondiente al número de vecinos a considerar para predecir la pertenencia al grupo de un punto de datos de prueba. En este trabajo, se obtuvo la mejor exactitud promedio al usar $k = 3$.

Con respecto a los modelos más complejos como SVM y RF, su optimización requirió de mayor consumo de tiempo. La optimización de los parámetros más importante para SVM incluyó la especificación de los parámetros de la función Kernel y el parámetro de costo (C). Se seleccionó la función de base radial (RBF) para la función Kernel del algoritmo. Se empleó el método de búsqueda Grid para la identificación de una combinación "óptima" de valores de parámetros al intentar diferentes valores y seleccionar aquellos que tienen el menor valor de error de entrenamiento. Para este trabajo los valores óptimos encontrados para los parámetros C y sigma del RBF, fueron de 4 y 0,02 respectivamente.

Finalmente el algoritmo RF contó con dos parámetros a optimizar, $mtry$ y nt , correspondientes a número total de variables a considerar para dividir en cada nodo (muestreo estratificado) y al número de árboles, respectivamente, calculados mediante el uso de la técnica de validación cruzada descrita anteriormente teniendo en cuenta la tasa

máxima de accuracy como criterio para su selección. El parámetro de ajuste 'mtry' se mantuvo constante en un valor de 5,29 mientras que para el número de árboles, el valor predeterminado fue de 500.

Una vez que se seleccionaron los valores óptimos para cada modelo, la sensibilidad (muestras pertenecientes a una clase y clasificadas correctamente en esta clase), especificidad (muestras que no pertenecen a una clase y clasificadas correctamente como no pertenecientes), y la tasa de exactitud o accuracy general, fueron los parámetros propuestos para la evaluación de los modelos de clasificación supervisados en el conjunto de datos de prueba o testeo. En la Tabla 4.8. se presentan los resultados que indican el rendimiento de los diferentes métodos de clasificación.

Tabla 4.8. Comparación de la performance de los diferentes métodos quimiométricos aplicados en las muestras de prueba.

Grupos	LDA		k-NN (k = 3)		SVM (C = 4; $\sigma = 0,02$)		RF (nt = 500; mtry = 5)	
	Sens. (%)	Espec. (%)	Sens. (%)	Espec. (%)	Sens. (%)	Espec. (%)	Sens. (%)	Espec. (%)
Mza	93	100	100	90	100	90	100	80
Noa	100	97	86	97	100	100	71	100
Ptg	100	97	67	100	67	100	100	100
Accuracy (%)	94,7		94,7		97,4		94,7	

A partir del análisis de la tabla 4.8, se observa que los cuatro métodos quimiométricos evaluados mostraron similares grados de éxito en la predicción de las muestras de testeo o prueba de vinos tintos. Los modelos LDA, k-NN y RF presentaron una igual tasa de accuracy, mientras que el método de SVM fue el que mostró el mejor resultado para la clasificación de las muestras de vino tinto según su procedencia, con un valor de accuracy del 97%. Por lo tanto, en base a los resultados, se obtuvo una muy buena discriminación de las muestras en estudio, que puede explicarse por la flexibilidad y la capacidad del algoritmo de SVM para crear un modelo generalizado, incluso para grupos de entrenamiento pequeños y no lineales gracias a su función Kernel.

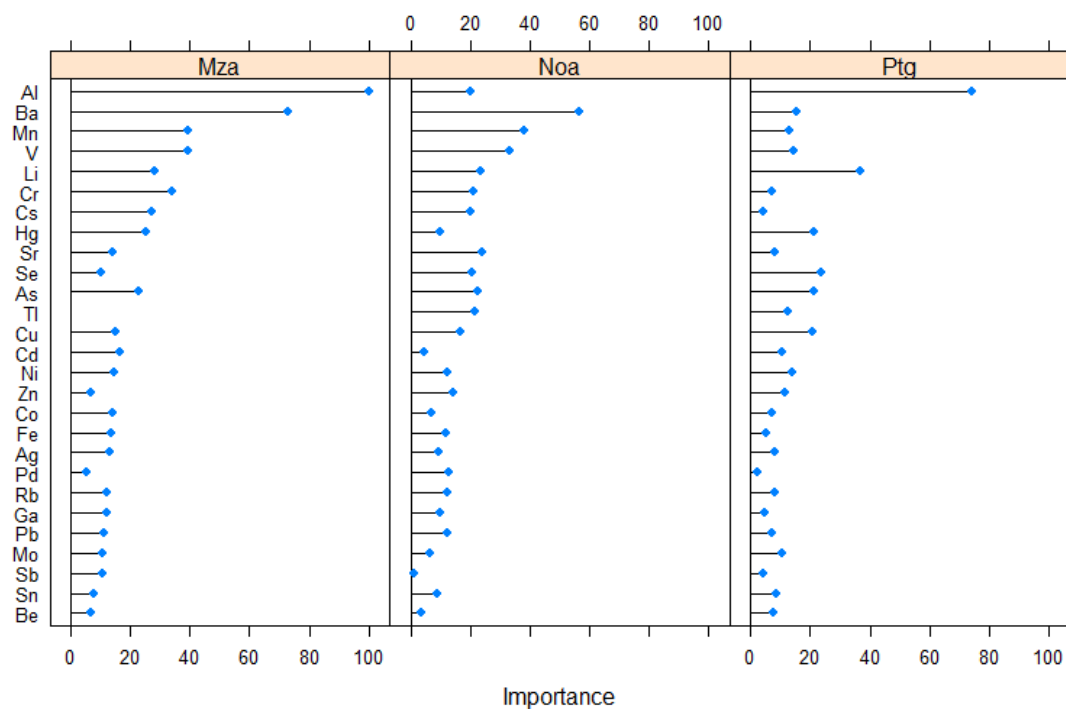


Figura 4.19. Importancia de las variables en RF según la región geográfica.

RF es un algoritmo que permite calcular la importancia de las variables (Random Forest Importance, RFI), obteniendo un ranking de los predictores, lo cual refleja la importancia de las variables en la predicción de la respuesta y, por lo tanto, la oportunidad de poder elegir las variables con la mejor capacidad predictiva. La Fig. 4.19 presenta un ranking de las variables para el método de clasificación de vino tinto aplicando RF en las muestras de testeo para cada

región de procedencia. Se puede observar que para el caso de la clasificación de vinos de Mendoza y Patagonia, Al tiene el mayor poder de discriminación, con una importancia del 100% y casi el 80% respectivamente, seguidos por Ba, Mn, V, Li y Cr en importancia para la clasificación geográfica de vinos de las tres regiones. Algunos de estos elementos se condice por lo expuesto por Phol y otros autores, quienes afirman que en particular las concentraciones de Ba, Ca, Cr, Li, Mg, Mn, Na, Rb y Sr son los reflejos más discriminatorios de la composición del suelo característico de una región geográfica distinta [30,31].

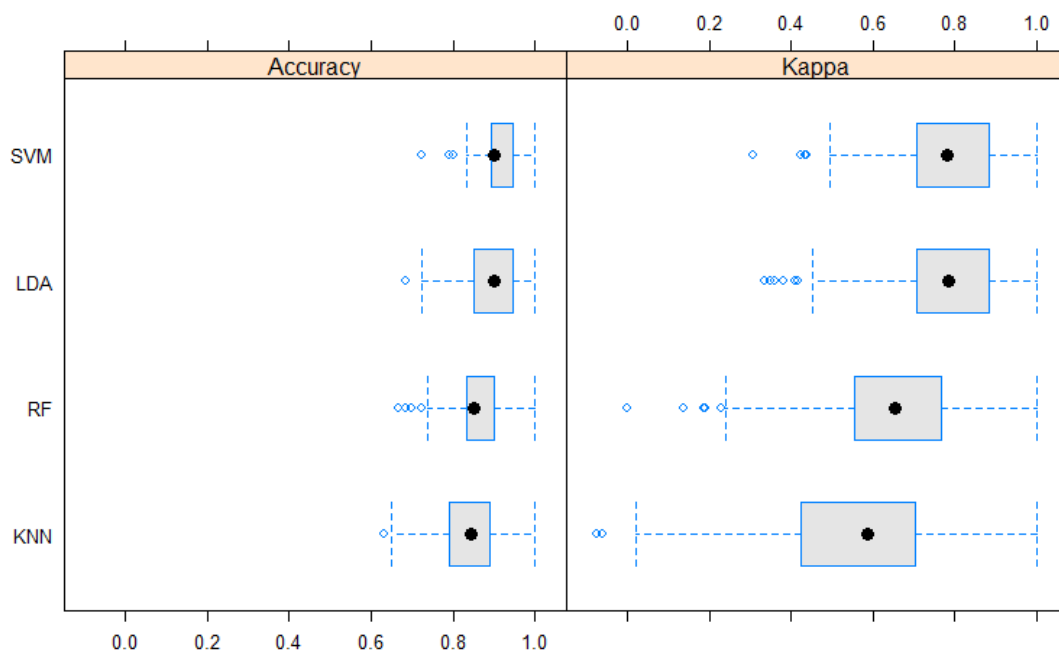


Figura 4.20. Gráficos box-plot comparando los modelos evaluados para la clasificación geográfica de vinos.

Finalmente, se realizó una comparación adicional del desempeño de los cuatro métodos de clasificación aplicados mediante el uso de la repetición del estimador por validación cruzada repetida 250 veces. Este procedimiento proporcionó una distribución de la accuracy general para comparar el rendimiento de cada método en su configuración de parámetros óptima. En la Fig. 4.20 se observa las distribuciones representadas por gráficos de caja y bigotes o box-plot. Los límites alrededor de la media representan la amplitud de la

dispersión de los resultados obtenidos para cada ciclo de remuestreo. Los resultados obtenidos por este enfoque confirman el excelente rendimiento de los cuatros modelos evaluados, mostrando una exactitud de alrededor del 90% para los métodos de LDA y SVM.

4.3.1.5. Conclusiones

La evaluación de la calidad y autenticidad del vino puede considerarse una demanda social, ya que los consumidores requieren productos de alta calidad. Para garantizar esto, la trazabilidad de la producción de vino, para preservar la calidad y el conocimiento de todo el proceso de elaboración del vino desde el suelo hasta la botella, se ha revelado como la mejor herramienta. Para ello, se deben aplicar herramientas analíticas químicas y quimiométricas para obtener información relacionada con un vino específico.

En este trabajo, la aplicación de la técnica de análisis multielemental ICP-MS, combinada con métodos quimiométricos, tales como LDA, k-NN, SVM y RF, proporcionaron herramientas robustas para la clasificación geográfica de vinos tintos elaborados en las tres regiones vitivinícolas más importantes de Argentina, Mendoza, el noroeste argentino y la Patagonia. Los resultados de la predicción demostraron que los cuatro modelos presentaron excelentes desempeños, siendo el modelo SVM propuesto, el método más eficaz y eficiente para la clasificación de las muestras de vino. Además, el modelo presentó un valor de costo o penalización de 4 y de sigma de la función Kernel de 0,02, seleccionados como los mejores parámetros para la clasificación de la muestra de vinos.

Por lo tanto, se puede considerar que los métodos quimiométricos combinados con el análisis multielemental son herramientas de gran utilidad para desarrollos en la clasificación y autenticidad del vino, lo que traería grandes beneficios para la industria y los laboratorios que trabajan en la producción de vinos de alta calidad y, en última instancia, promovería la aceptación por parte de los consumidores.

4.3.1.6. Referencias

- [1] International Organization of Vine and Wine (OIV), International Code of Oenological Practices, 33 (2016) 1–29. <http://www.oiv.int/en/technical-standards-and-documents/oenological-practices/international-code-of-oenological-practices> (accessed September 29, 2019).
- [2] C. Villano, M.T. Lisanti, A. Gambuti, R. Vecchio, L. Moio, L. Frusciante, R. Aversano, D. Carputo, Wine varietal authentication based on phenolics, volatiles and DNA markers: State of the art, perspectives and drawbacks, *Food Control*. 80 (2017) 1–10. doi:10.1016/j.foodcont.2017.04.020.
- [3] G. Grindlay, J. Mora, L. Gras, M.T.C. de Loos-Vollebregt, Atomic spectrometry methods for wine analysis: A critical evaluation and discussion of recent applications, *Anal. Chim. Acta*. 691 (2011) 18–32. doi:10.1016/j.aca.2011.02.050.
- [4] International Organization of Vine and Wine (OIV), State of the world vitivinicultural sector in 2019, (2019) 15. <http://www.oiv.int/public/medias/7298/oiv-state-of-the-vitivinicultural-sector-in-2019.pdf>.
- [5] M.E. Alañón, M.S. Pérez-Coello, M.L. Marina, Wine science in the metabolomics era, *TrAC Trends Anal. Chem.* 74 (2015) 1–20. doi:10.1016/j.trac.2015.05.006.
- [6] C.A. Teixeira dos Santos, R.N.M.J. Páscoa, M.C. Sarraguça, P.A.L.S. Porto, A.L. Cerdeira, J.M. González-Sáiz, C. Pizarro, J.A. Lopes, Merging vibrational spectroscopic data for wine classification according to the geographic origin, *Food Res. Int.* (2017). doi:10.1016/j.foodres.2017.09.018.
- [7] L. Pereira, S. Gomes, S. Barrias, E. Gomes, M. Baleiras-Couto, J. Fernandes, P. Martins-Lopes, From the Field to the Bottle—An Integrated Strategy for Wine Authenticity, *Beverages*. 4 (2018) 71. doi:10.3390/beverages4040071.
- [8] C. van Leeuwen, 9 – Terroir: the effect of the physical environment on vine growth, grape ripening and wine sensory attributes, in: *Manag. Wine Qual.*, 2010: pp. 273–315.

- doi:10.1533/9781845699284.3.273.
- [9] C. Zucca, R. Biancalani, S. Kapur, E. Akça, P. Zdruli, L. Montanarella, F. Nachtergaele, *Soil Security for Ecosystem Management*, 8 (2014) 31–59. doi:10.1007/978-3-319-00699-4.
- [10] F. Foroni, M. Vignando, M. Aiello, V. Parma, M.G. Paoletti, A. Squartini, R.I. Rumiati, The smell of terroir! Olfactory discrimination between wines of different grape variety and different terroir, *Food Qual. Prefer.* 58 (2017) 18–23. doi:10.1016/j.foodqual.2016.12.012.
- [11] J. Saurina, Characterization of wines using compositional profiles and chemometrics, *TrAC Trends Anal. Chem.* 29 (2010) 234–245. doi:10.1016/j.trac.2009.11.008.
- [12] The European Parliament and the Council of the European Union, REGULATION (EU) No 1151/2012 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 21 November 2012 on quality schemes for agricultural products and foodstuffs, *Off. J. Eur. Union.* L343 (2012) 1–29.
- [13] M. Kamal, R. Karoui, Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review, *Trends Food Sci. Technol.* 46 (2015) 1–22. doi:10.1016/j.tifs.2015.07.007.
- [14] G.P. Danezis, A.S. Tsagkaris, F. Camin, V. Brusic, C.A. Georgiou, Food authentication: Techniques, trends & emerging approaches, *TrAC - Trends Anal. Chem.* 85 (2016) 123–132. doi:10.1016/j.trac.2016.02.026.
- [15] M. Likar, K. Vogel-Mikuš, M. Potisek, K. Hančević, T. Radić, M. Nečemer, M. Regvar, Importance of soil and vineyard management in the determination of grapevine mineral composition, *Sci. Total Environ.* 505 (2015) 724–731. doi:10.1016/j.scitotenv.2014.10.057.
- [16] D. Cozzolino, S. Heather, *Analytical and chemometric-based methods to monitor and evaluate wine protected designation*, 1st ed., Copyright © 2013 Elsevier B.V. All rights reserved., 2013. doi:10.1016/B978-0-444-59562-1.00015-3.

- [17] R.D. Di Paola-Naranjo, M. V. Baroni, N.S. Podio, H.R. Rubinstein, M.P. Fabani, R.G. Badini, M. Inga, H.A. Oстера, M. Cagnoni, E. Gallegos, E. Gautier, P. Peral-García, J. Hoogewerff, D.A. Wunderlin, Fingerprints for Main Varieties of Argentinean Wines : Terroir Differentiation by Inorganic , Organic , and Stable Isotopic Analyses Coupled to Chemometrics, (2011) 7854–7865.
- [18] H. Shimizu, F. Akamatsu, A. Kamada, K. Koyama, K. Iwashita, N. Goto-Yamamoto, Variation in the mineral composition of wine produced using different winemaking techniques, *J. Biosci. Bioeng.* (2020). doi:10.1016/j.jbiosc.2020.03.012.
- [19] H. Shimizu, F. Akamatsu, A. Kamada, K. Koyama, M. Okuda, H. Fukuda, K. Iwashita, N. Goto-Yamamoto, Discrimination of wine from grape cultivated in Japan, imported wine, and others by multi-elemental analysis, *J. Biosci. Bioeng.* 125 (2018) 413–418. doi:10.1016/j.jbiosc.2017.10.016.
- [20] M. Aceto, The Use of ICP-MS in Food Traceability, in: *Adv. Food Traceability Tech. Technol.*, Elsevier, 2016: pp. 137–164. doi:10.1016/B978-0-08-100310-7.00008-9.
- [21] E.P. Pérez-Álvarez, R. Garcia, P. Barrulas, C. Dias, M.J. Cabrita, T. Garde-Cerdán, Classification of wines according to several factors by ICP-MS multi-element analysis, *Food Chem.* 270 (2019) 273–280. doi:10.1016/j.foodchem.2018.07.087.
- [22] M.J. Cabrita, N. Martins, P. Barrulas, R. Garcia, C.B. Dias, E.P. Pérez-Álvarez, A.M. Costa Freitas, T. Garde-Cerdán, Multi-element composition of red, white and palhete amphora wines from Alentejo by ICPMS, *Food Control.* 92 (2018) 80–85. doi:10.1016/j.foodcont.2018.04.041.
- [23] F. Soares, M.J. Anzanello, F.S. Fogliatto, M.C.A. Marcelo, M.F. Ferrão, V. Manfroi, D. Pozebon, Element selection and concentration analysis for classifying South America wine samples according to the country of origin, *Comput. Electron. Agric.* 150 (2018) 33–40. doi:10.1016/j.compag.2018.03.027.
- [24] B. Tariba, Metals in Wine—Impact on Wine Quality and Health Outcomes, *Biol. Trace*

- Elem. Res. 144 (2011) 143–156. doi:10.1007/s12011-011-9052-7.
- [25] S. Rocha, E. Pinto, A. Almeida, E. Fernandes, Multi-elemental analysis as a tool for characterization and differentiation of Portuguese wines according to their Protected Geographical Indication, *Food Control*. 103 (2019) 27–35. doi:10.1016/j.foodcont.2019.03.034.
- [26] S. Orellana, A.M. Johansen, C. Gazis, Geographic classification of U.S. Washington State wines using elemental and water isotope composition, *Food Chem. X*. 1 (2019) 100007. doi:10.1016/j.fochx.2019.100007.
- [27] P. Pohl, What do metals tell us about wine?, *TrAC - Trends Anal. Chem.* 26 (2007) 941–949. doi:10.1016/j.trac.2007.07.005.
- [28] (OIV) International Organization of Vine and Wine, Maximum acceptable limits of various substances contained in wine. Code International des Pratiques Oenologiques. Fiche Code OIV, 2015.
- [29] R.G. Brereton, Pattern recognition in chemometrics, *Chemom. Intell. Lab. Syst.* 149 (2015) 90–96. doi:10.1016/j.chemolab.2015.06.012.
- [30] P. Pohl, What do metals tell us about wine?, *TrAC - Trends Anal. Chem.* 26 (2007) 941–949. doi:10.1016/j.trac.2007.07.005.
- [31] M. Koreňovská, M. Suhaj, Identification of some Slovakian and European wines origin by the use of factor analysis of elemental data, *Eur. Food Res. Technol.* 221 (2005) 550–558. doi:10.1007/s00217-005-1193-5.

CAPÍTULO V:
CONCLUSIONES Y
PERSPECTIVAS FUTURAS

5.1. CONCLUSIONES Y PERSPECTIVAS FUTURAS

La calidad y autenticidad de los productos vitivinícolas relacionada con el origen geográfico, son atributos de suma importancia especialmente en el mundo del vino de hoy en día, ya que los consumidores exigen productos de alta calidad y los productores posicionan sus productos en el mercado brindando un valor agregado y por ende mayor distinción a los mismos teniendo en cuenta su procedencia. Para garantizar esto, la trazabilidad surge como un procedimiento relevante, basado en sistemas de denominación de origen, con reglamentaciones claramente definidas y establecidas no sólo a nivel país, sino también en todo el mundo.

La química de la uva y sus derivados, y las diversas características que influyen en su composición, convierten este tipo de muestras en matrices complejas para analizar y garantizar su autenticidad y calidad. Sin embargo, la existencia de técnicas instrumentales capaces de generar información múltiple, puede ser un gran aliado cuando se combina con herramientas quimiométricas multivariadas, para resolver las necesidades de establecer la identidad y clasificación de las uvas y sus productos derivados.

En el presente trabajo de tesis doctoral, a partir del empleo de la técnica de ICP-MS, se determinaron las concentraciones de diferentes elementos de la tabla periódica en muestras de semillas y hollejos de uva y productos derivados como jugo de uva y vino procedentes de la provincia de Mendoza y de diversas regiones de importancia vitivinícola de Argentina, con el objetivo de encontrar posibles marcadores químicos de identidad que permitieran discriminar estos productos en función de su origen geográfico.

A partir de los datos de concentraciones obtenidas en los distintos tipos de muestras analizadas, se procedió a un análisis exploratorio que incluyó la presentación de los datos en gráficos tipo de caja y bigotes o box-plot, permitiendo observar las distribuciones de los valores de concentraciones en las muestras bajo estudio. Posteriormente el análisis de componentes principales (PCA) posibilitó la determinación de las principales fuentes de

variabilidad en el conjunto de datos, y la visualización de la agrupación natural de los objetos estudiados.

Los resultados de PCA mostraron para el caso de las muestras de semillas de uva, que la primera componente principal (PC1) representó el 53,6% de la varianza total y la siguiente componente principal, 10,6% (PC2). A pesar de estos altos valores de variabilidad explicada por las dos primeras componentes principales, en el gráfico de scores se observó una gran superposición entre las muestras de diferentes orígenes, sólo las muestras de dos departamentos pudieron ser diferenciadas del resto de los grupos de otras procedencias. Para las muestras de hollejos de uva, los resultados mostraron que las dos primeras componentes explicaron el 34% de la varianza total, la primera componente principal representó el 23,9%, requiriendo casi 12 PC para explicar el 80% de la varianza. Estos resultados se vieron plasmados en el gráfico de las observaciones o scores, en el cual se pudo también observar una gran superposición entre las muestras, no distinguiéndose una separación de acuerdo al origen geográfico de las mismas. En el caso de las muestras de jugos de uva, la aplicación de PCA permitió una reducción de dos primeras componentes principales, explicando el 53,6% de la varianza total. En el gráfico de scores se puede observar una buena separación entre las muestras de jugos de Argentina y Brasil, esto se debió quizás a la mayor distancia del origen de las muestras bajo estudio. Por último, para el caso de las muestras de vino tinto, con el análisis de PCA las primeras dos componentes principales (PC) representaron casi el 30% de la variabilidad. En el gráfico de observaciones de las dos primeras componentes principales, se observó una superposición entre las puntuaciones correspondientes a las diferentes muestras identificadas según su origen.

Una vez realizado el análisis exploratorio de los datos, y ante los resultados no satisfactorios del análisis de PCA que permitieran una diferenciación de las muestras en base a su origen, en las muestras de semillas, hollejos y vinos; aunque no tanto en las muestras de jugos de uva, se procedió a evaluar diferentes modelos quimiométricos clasificatorios más

avanzados. Para ello, se propusieron siete diferentes tipos de algoritmos con distintos grados de complejidad: análisis discriminante lineal (LDA), análisis discriminante de mínimos cuadrados parciales (PLS-DA), regresión logística multinomial (MLR), k- vecinos más cercanos (k-NN), máquina de vectores de soporte (SVM), redes neuronales (NNET) y bosques aleatorios (RF). Estos algoritmos se utilizaron teniendo en cuenta las características de cada tipo de matriz de datos de las muestras bajo estudio.

Los resultados obtenidos en la clasificación geográfica de los distintos tipos de muestras mostraron en todos los casos, mejores valores en la performance de aquellos algoritmos que tienen una función no lineal como límite de decisión. En el caso de las muestras de semillas y hollejos de uva, el algoritmo RF obtuvo el mayor valor de exactitud en la clasificación, con valores de casi el 100% y 90% respectivamente. El modelo propuesto por el algoritmo SVM, presentó también muy buenos resultados de exactitud para la discriminación de estas muestras según su procedencia. Estos resultados evidencian el potencial de ambos algoritmos para discriminar las semillas y hollejos de uvas cultivadas en departamentos cercanos dentro de la provincia de Mendoza, donde se podrían esperar no grandes diferencias en la composición del suelo y por ende en la química de las uvas. En el caso de los jugos de uva, al contar con muestras procedentes de Argentina y Brasil, los métodos propuestos para realizar la clasificación geográfica, presentaron excelentes resultados de exactitud, con valores alrededor del 90%. Las muestras de vino tinto procedentes de diferentes regiones vitivinícolas de Argentina, se lograron diferenciar geográficamente con los cuatro algoritmos evaluados, presentando SVM el mayor valor de exactitud, con un valor de 97,4%.

En base a lo anteriormente expuesto, se puede concluir que la quimiometría se presenta como una herramienta efectiva para el análisis rápido y económico de los resultados experimentales, que permite la clasificación de uvas y productos derivados según su origen geográfico. Combinando sus algoritmos con técnicas analíticas de análisis multielemental, como el ICP-MS, surgen herramientas poderosas y robustas para la toma de huellas dactilares

o “fingerprint” de las uvas y sus productos derivados, considerando a los perfiles minerales como excelentes marcadores del origen. Por otro lado, es importante mencionar que la necesidad de grandes bases de datos construidas con la composición química obtenida de numerosas muestras, puede considerarse una de las principales desventajas de estas herramientas quimiométricas.

Esta tesis presenta resultados que demuestran que las herramientas quimiométricas de clasificación pueden considerarse muy útiles para garantizar la trazabilidad y autenticidad de la uva y sus productos derivados. Aunque sea necesario el empleo de técnicas instrumentales capaces de recopilar gran cantidad de información sobre las muestras, el uso de estas técnicas analíticas en estrecha relación con las herramientas quimiométricas a través de software especialmente diseñados para integrar las rutinas operativas con cálculos matemáticos y estadísticos complejos, junto con su interpretación rápida y automatizada, traería grandes beneficios para la industria y los laboratorios que trabajan en el control rutinario de las materias primas y elaboración de vinos. Todo ello resultando en la producción de vinos y demás productos de alta calidad que promovería la aceptación por parte de los consumidores y expandiría el comercio.