



UNIVERSIDAD DE SEVILLA

FACULTAD DE QUÍMICA

DEPARTAMENTO DE QUÍMICA ANALÍTICA

**CARACTERIZACIÓN ANALÍTICA Y
DIFERENCIACIÓN GEOGRÁFICA DE PIMENTÓN
MEDIANTE TÉCNICAS DE RECONOCIMIENTO
DE PATRONES**

TESIS DOCTORAL

ANA PALACIOS MORILLO

Sevilla, 2015

D. Agustín García Asuero, Catedrático Director del Departamento de Química Analítica de la Universidad de Sevilla.

CERTIFICA: Que el presente trabajo de investigación ha sido realizado en los laboratorios del Departamento de Química Analítica de la Universidad de Sevilla y reúne las condiciones exigidas a los trabajos de tesis doctoral.

Y para que conste, expido y firmo el presente certificado en Sevilla a 3 de Septiembre de 2015.

A circular blue stamp of the Department of Analytical Chemistry, University of Seville, is partially obscured by a large, stylized handwritten signature in blue ink. The stamp contains the text 'DEPARTAMENTO DE QUÍMICA ANALÍTICA - SEVILLA' and 'UNIVERSIDAD DE SEVILLA' around a central emblem.

Fdo. Dr. D. Agustín García Asuero

DIRECTORES



Fdo. D. Fernando de Pablos Pons
Catedrático de Universidad
Departamento de Química Analítica
Universidad de Sevilla



Fdo.D. José Marcos Jurado Jurado
Profesor Titular de Universidad
Departamento de Química Analítica
Universidad de Sevilla

Memoria presentada para aspirar al grado de Doctor en Ciencias Químicas.



Fdo. Ana Palacios Morillo

Ldo. Química

Quiero expresar mi más sentido agradecimiento:

Al Dr. Agustín García Asuero, Director del Departamento de Química Analítica de la Universidad de Sevilla, por haberme dado la oportunidad de realizar este trabajo en dicho departamento

Al personal de administración y servicios del Departamento de Química Analítica de la Universidad de Sevilla, Sofía, Charo, Joaquín y Gustavo por haber estado siempre dispuestos a prestar su ayuda en lo que he necesitado.

Al personal de administración y servicios de la Facultad de Química, al personal de la Conserjería, Javier, Rocho, M^o del Mar, por hacer que la Facultad se sienta como una casa. A Julio Carpio y Paco Hernández por haberme hecho tan fácil unos comienzos tan difíciles para mí. A mi equipo de desayuno Jaime, Javi y sobretodo Jesús por hacer de las mañanas un ratito estupendo.

A mis compañeros de Química Inorgánica, a José Antonio Odriozola por tenerme presente como a una más de su equipo, a Paqui por tener siempre un plan de futuro preparado para mí, a Leo por los consejos que me ha dado en todas las tardes de prácticas compartidas, a mi segundo padre Agustín Galindo por aguatarme los llantos de agobio desde hace por lo menos ocho años. A Juan y Jesús. A Pilar Malet y Rafael Andreu por el esfuerzo y trabajo que ponen a todo lo que hacen y el cariño que transmiten. Y por supuesto a mi Svet y mi Maribel por ser compañeras y equipo OBL, además de amigas.

A Gustavo González, por enseñarme muchas cosas de estadística, por prestarme sus cuadernillos de problemas y por todas esas charlas de tarde que hemos tenido.

A mis amigas de la facultad, Bea, Mopa, Purga y Xusi, por todos los buenos momentos que hemos pasado en clase y por el pasillo comiendo. A Samuel por estar siempre ahí. A las viejas del Visillo, a Belén y mi clan de Huelva por entender que no siempre he podido estar cerca. A Álvaro e Irene por hacer más fáciles los cambios y las cosas nuevas. A Pepe, Raúl, Koki, MJ, Don Ángel y Santana por esos viajes que nos han dado fuerza para volver a septiembre.

A Ángela por plantear los problemas desde un enfoque diferente y dar esa perspectiva que no siempre tengo, además de la ayuda en el laboratorio. A Daniel también por el tiempo que ha invertido aquí con nosotros.

A Ara que es capaz de hacer feliz a cualquiera que se ponga a su lado incluida a mí, por quererme tanto aunque la llame Araceli.

A mi familia Peruana, Hermano Nino, Blanquita y Hermano Félix por acogernos en vuestra casa y hacer que seamos un poco más felices y darnos otra visión del mundo.

A Tomás, por haber compartido el principio de todo esto juntos, por haberme aconsejado y aguantado en los momentos de agobio. Por ser el mejor doctorado del mundo (con premio incluido) y seguir siendo el mismo coleta de siempre.

A Kiko por ser el mejor amigo y compañero de piso del mundo, por estar para lo bueno y para lo malo sin importar hora, día mes o año. A Carmen, por haberme aguantado estar contenta, triste, agobiada, tranquila, por no tener ni idea de química y aun así que pareciera que te interesaba. Por aconsejar mi *outfit* mi primer día de profesora, por todo, gracias.

A mis directores de tesis. Fernando, porque desde antes de entrar en la facultad ya estaba pendiente de que estuviera bien, por querer que entrara en el grupo de investigación, por sus consejos siempre orientados hacia mi bien, por quitarle importancia a cuando lloro y por la tranquilidad que da saber que hay alguien cerca en quien confiar. A Marcos, porque aparte de director ha sido profesor, compañero de almuerzo (tapa/taper) y amigo. Por enseñarme todo lo sé de quimiometría y por entender siempre mis decisiones aunque no fueran compartidas.

A mis padres y mi hermano. A mi padre, que además de agradecerle como Cela le agradeció a su madre, ha estado siempre pendiente y preocupado por mí y mi trabajo a pesar de que no siempre compartimos ideas. ¡Ah! Y por las quince millones de veces que me ha preguntado *Ali, ¿Cuándo vas a leer la tesis?*. A mi madre porque siempre tiene las palabras mágicas para todos los momentos, desde *venga Ana un esfuercito más, eso te pasa por dejarlo para el final o haz lo que creas que está bien que yo te apoyo*. Y a mi hermano porque sí, por alegrar la casa desde que nació.

Y como no a Pablo, por haberme ayudado con esto desde los agradecimientos hasta la bibliografía, por haberse leído y releído 500 páginas buscando comas y puntos mal puestos, numeraciones raras o figuras descolocadas, por plantear matricularse de Química en la UNED para poder ayudar todavía más, pero sobre todo por estar a mi lado, porque es quien más ha sufrido esto sin lugar a dudas, y aun así no se ha quejado ni un segundo. Por aguantar mi buen humor y mi mal humor, y por dar a mi vida un sentido perfecto y por ser él, perfecto.

INDICE

Abreviaturas

Objetivos

I. Introducción	1
I.1. Aspectos Históricos	1
I.1.1. Pimiento y pimentón	1
I.1.2. Entrada del pimiento en Europa	3
I.2. Importancia del Género <i>Capsicum</i>	3
I.3. Pimentón. Definición, elaboración y tipos.	8
I.3.1. Definición y elaboración	8
I.3.2. Denominaciones de origen de pimentón en España	9
I.3.2.1. Pimentón de La Vera (Extremadura)	9
I.3.2.2. Pimentón de Murcia (Murcia)	12
I.4. Parámetros de calidad del pimentón	13
I.5. Composición mineral y orgánica	14
I.5.1. Métodos de análisis de compuestos orgánicos en pimientos y pimentón	17
I.5.1.1. Determinación mediante cromatografía líquida de alta resolución	18
I.5.1.2. Determinación mediante cromatografía de gases	21
I.5.2. Determinación de metales en pimiento y pimentón	23
I.5.3. Otras técnicas usadas en el análisis de pimiento y pimentón	24
I.6. Aplicación de métodos de reconocimiento de patrones a la caracterización de alimentos.	26
I.7. El color	33
I.7.1. Breve introducción histórica	33
I.7.2. Definición del color	35
I.7.3. Medidas del color	36
I.7.4. Ordenación del color	39
I.7.5. Espacios de color	43
I.7.5.1. Espacio CIELab	44
I.7.6. El color en pimientos y pimentones	47
I.8. Bibliografía	51

II. Quimiometría	75
II.1 Introducción a la quimiometría	75
II.2. Pruebas de normalidad	77
II.3. Contrastes de significación paramétricos	81
II.3.1. Contraste t de Student	81
II.3.2. Análisis de la varianza	83
II.3.2.1. ANOVA de un factor	83
II.3.2.2. ANOVA de dos factores anidados	85
II.3.2.3. Prueba de la genuina diferencia significativa de Tukey	88
II.4. Contrastes de significación no paramétricos	90
II.4.1. Prueba U de Mann-Whitney	90
II.4.2. Test de Kruskal- Wallis	91
II.5. Detección de valores extremos	92
II.5.1. Valores extremos: definición y tipos	92
II.5.2. Detección de outliers centrales	93
II.5.3. Detección de outliers dispersivos	94
II.6. Diseño experimental	95
II.6.1. Diseño experimental simple	97
II.6.2. Diseño experimental factorial	97
II.6.2.1. Diseño factorial a dos niveles	97
II.6.2.2. Diseño factorial a más de dos niveles	100
II.6.3. Diseño experimental de bloque y fraccional	100
II.6.4. Evaluación de diseños experimentales: función deseada	102
II.7. Reconocimiento de patrones	104
II.7.1. Introducción a los métodos de reconocimiento de patrones	104
II.7.1.1. Conceptos básicos	106
II.7.1.2. Procesado de datos	110
II.7.2. Análisis de componentes principales	113
II.7.3. Reconocimiento de patrones no supervisado:	118
II.7.4. Reconocimiento de patrones supervisado	118
II.7.4.1. Análisis discriminante lineal	120
II.7.4.2. Máquinas de vectores soporte	127

II.7.4.3. Modelado suave independiente por analogía de clases	134
II.8. Regresión múltiple y multivariante	136
II.8.1. Modelos de regresión	136
II.8.2. Regresión en componentes principales	139
II.8.3. Regresión por mínimos cuadrados parciales	139
II.9. Reconocimiento de patrones mediante algoritmos basados en redes neuronales artificiales	140
II.9.1. Tipos y clasificaciones de las redes neuronales artificiales	142
II.9.2. Neuronas y redes	144
II.9.3. Aprendizaje no supervisado	149
II.9.4. Aprendizaje no supervisado	149
II.10. Bibliografía	152
III. Parte Experimental	161
III.1. Material y Reactivos	161
III.1.1. Aparatos	161
III.1.2. Reactivos y disoluciones	161
III.1.2.1. Reactivos	162
III.1.2.2. Disoluciones estándar	162
III.1.3. Muestras	163
III.2. Procedimientos	169
III.2.1. Determinación de la humedad en muestras de pimentón	169
III.2.2. Determinación de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn mediante ICP-AES	169
III.2.3. Medidas espectrofotométricas	171
III.2.3.1. Determinación del parámetro ASTA	172
III.2.3.2. Registro de espectros para el cálculo de las coordenadas CIELab	172
III.2.3.3. Registro de espectros para el cálculo del parámetro SHU	172
III.2.4. Determinación de capsaicinoides mediante HPLC-UV	173
III.3. Bibliografía	174
IV. Discusión y resultados	175
IV.1. Determinación de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn por ICP-AES	175

IV.1.1.Comparación de los métodos de mineralización	175
IV.1.2.Efecto Matriz	190
IV.1.3. Veracidad	200
IV.1.4. Precisión	200
IV.1.5. Calibración y linealidad	203
IV.1.6. Límite de detección y límite de cuantificación	203
IV.1.7. Determinación de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn en muestras de pimentón	205
IV.2. Diferenciación geográfica de pimentón según su perfil metálico	207
IV.2.1. Análisis estadístico básico	207
IV.2.2. Aplicación de técnicas de reconocimiento de patrones	220
IV.2.2.1. Diferenciación geográfica de pimentones dulces	221
IV.2.2.2 Diferenciación geográfica de pimentones picantes	231
IV.2.2 3. Diferenciación geográfica de pimentones agridulces	240
IV.2.2.4. Diferenciación geográfica mediante un modelo global	247
IV.3. Diferenciación geográfica de pimentón mediante el estudio del color y los espectros de absorbancia ultravioleta visible	256
IV.3.1. Estudio del color en base al parámetro ASTA y las coordenadas CIELab para la diferenciación de pimentones	257
IV.3.1.1. Diferenciación geográfica de muestras de pimentones dulces a partir del parámetro ASTA y coordenadas CIELab	267
IV.3.1.2 Diferenciación geográfica de muestras de pimentones picantes a partir del parámetro ASTA y coordenadas CIELab	270
IV.3.1.3. Diferenciación geográfica de muestras de pimentones agridulces a partir del parámetro ASTA y coordenadas CIELab	274
IV.3.2. Estudio del color en base al espectro ultravioleta visible.	277
IV.3.2.1. Diferenciación geográfica de muestras de pimentones dulces a partir del espectro ultravioleta visible.	278
IV.3.2.2. Diferenciación geográfica de muestras de pimentones picantes a partir del espectro ultravioleta visible.	283
IV.3.2.3. Diferenciación geográfica de muestras de pimentones agridulces a partir del espectro ultravioleta visible.	289
IV.3.3. Comparación de modelos de clasificación	295

IV.4. Determinación de capsaicinoides en pimentón mediante cromatografía líquida de alta eficacia.	303
IV.4.1. Optimización de las condiciones cromatográficas	304
IV.4.1.1. Asimetría y factor de cola	312
IV.4.2. Optimización de las condiciones de extracción	313
IV.4.3. Efecto matriz	318
IV.4.4 Parámetros de calidad del método	320
IV.4.4.1. Veracidad	320
IV.4.4.2. Precisión	321
IV.4.4.3. Calibración, linealidad y límites de detección y cuantificación	321
IV.4.5. Determinación de capsaicinoides en pimentones picantes	324
IV.4.6. Determinación de capsaicinoides en pimentones agridulces	328
IV.4.7. Determinación de capsaicinoides en pimentones dulces	331
IV.4.8. Modelos para predecir las unidades SHU en pimentones picantes a partir del espectro de UV-vis.	332
IV.5. Bibliografía	344
V. Resumen y Conclusiones	345
Anexos	357

Abreviaturas

AAS	Atomic Absorption Spectrometry
AM	Análisis Multivariante
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
ASTA	American Spice Trade Association
BOE	Boletín Oficial del Estado
CA	Cluster Analysis
CART	Classification and Regression Trees
CE	Captura Electrónica
CEE	Comunidad Económica Europea
CIE	Comission Internationale de L'Eclairage
CLS	Capsaicinoid-like substances
DAD	Diode Array Detector
DAGJK	Delete A Group Jackknife
DIN	Deutsches Institut für Normung
DOP	Denominación de Origen Protegida
FAAS	Flame Atomic Absorption Spectroscopy
FAES	Flame Atomic Emission Spectroscopy
FAO	Food and Agriculture Organization of the United Nations
FID	Flame Ionisation Detector
FT	Fourier-Raman Transformed
GC	Gas Chromatography
HCA	Hierarchical Cluster Analysis
HG-AAS	Hydride Generation Atomic Absorption Spectroscopy
HPCE	High Performance Capillary Electrophoresis
HPLC	High Performance Liquid Chromatography
HRGC	High Resolution Gas Chromatography
HSL	Hue-Saturation-Luminosity
HSV	Hue-Saturation-Value
ICP-MS	Inductively Coupled Plasma Mass-Spectroscopy
ICP-AES	Inductively Coupled Plasma Emission Spectroscopy
IGP	Indicación Geográfica Protegida
IR	Infrared

KNN	K Nearest Neighbours
LC	Liquid Chromatography
LDA	Linear Discriminant Analysis
LOD	Limit Of Detection
LS-SVM	Least Square Support Vector Machine
MIR	Medium Infrared
MS	Mass Spectrometry
MW-PLS-DA	Moving Window Partial Least Square Discriminant Analysis)
NCS	Natural Colour System
NIPALS	Nonlinear Iterative Partial Least Square
NIR	Near Infrared
NPD	Nitrogen Phosphorous Detector
OPS	Ordered Prediction Selection
PCA	Principal Component Analysis
PCR	Principal Component Regression)
PLS	Partial Least Squares
PLS-DA	Partial Least Square Discriminant Analysis
PLSR	Partial Least Square Regression
PNN	Probabilistic Neuronal Network
PR	Pattern Recognition
PRESS	Predicted Residual Sum of Squares
RGB	Red Green Blue
RLS	Regresión Lineal Simple
RLM	Regresión Lineal Múltiple
RMN	Resonancia Magnética Nuclear
SHU	Scoville Heat Units
SIMCA	Soft Independent Modelling of Class Analogy
SPE	Solid Phase Extraction
SVM	Support Vector Machine
SWIR	Short Wave Infrared
TLC	Thin Layer Chromatography
TOF	Time of Flight
UE	Unión Europea
UNE	Una Norma Española

USDA	United States Department of Agriculture
UV-Vis	Ultraviolet-Visible
UV-NIR	Ultraviolet-Near Infrared
VIF	Variance Inflation Factor
VIS-NIR	Visible Near Infrared

Objetivos

El objetivo principal del presente trabajo es el desarrollo de métodos analíticos que permitan la diferenciación de manera eficaz y rápida de las dos denominaciones de origen protegidas de pimentón que existen en España. Para ello se han considerado diversos enfoques. Por una lado se tiene en cuenta el perfil metálico, y por otro, los parámetros del color.

Dado que no existe en la legislación un método para la determinación de metales en muestras de pimentón, ni material de referencia para tales análisis, uno de los objetivos del trabajo es proponer un método para la mineralización y posterior determinación de metales en muestras de pimentón de Extremadura y Murcia. Se consideran los elementos aluminio, boro, calcio, cobre, hierro, potasio, magnesio, manganeso, sodio, níquel, fósforo, plomo, estroncio y zinc. Debido a que los métodos de digestión son una etapa crítica en el método analítico, ya que son una fuente potencial de errores, se ha hecho especial hincapié en el estudio de esta etapa de análisis, probando diversas mezclas ácidas para las digestiones.

La etiqueta de denominación de origen implica cierta calidad en los productos, y por tanto una repercusión importante en el coste de estos. Por ello, otro objetivo de este trabajo es implementar una técnica sencilla, rápida y económica como es la espectroscopía ultravioleta visible, como herramienta analítica a utilizar en el proceso de diferenciación de orígenes de pimentón.

El carácter picante del pimentón es una de sus características más importantes. La medida de este carácter se realiza habitualmente a partir de la determinación de la concentración de capsaicinoides, que requiere una etapa previa de extracción y limpieza de la muestra, y posterior determinación, generalmente, mediante cromatografía líquida. Por tanto, otro objetivo de este trabajo ha sido realizar un estudio que permita simplificar la determinación del nivel de picante en muestras de pimentón. Para ello, se han considerado las posibilidades de la espectrofotometría ultravioleta visible.

Hasta la fecha no existen demasiados trabajos donde se apliquen técnicas de análisis multivariante y de reconocimiento de patrones para la caracterización de pimentones de orígenes españoles. Por ello, otro objetivo que se ha planteado en el

presente trabajo a sido utilizar el perfil metálico, los espectros ultravioleta visible, las coordenadas L^* , a^* , b^* y el parámetro ASTA, calculados a partir de espectros ultravioleta visible, como descriptores químicos que permitan la diferenciación de muestras de pimentón pertenecientes a las dos denominaciones de origen protegidas existentes en España. En primer lugar se evaluará la capacidad discriminatoria de los parámetros determinados con este fin. Posteriormente se tratará de establecer, mediante técnicas de reconocimiento de patrones, las reglas de clasificación que discriminen las muestras.

I. Introducción

I.1. Aspectos históricos

I.1.1. Pimiento y pimentón

El origen del descubrimiento del pimiento y el pimentón está estrechamente ligado con el descubrimiento de América, los viajes de Cristóbal Colón y el importante papel económico de la comercialización de lo que denominaban especias exóticas en la Europa medieval. Cristóbal Colón aseguró ser capaz de llegar a la *India del Lejano Oriente*, con la intención de encontrar la pimienta negra (*Piper nigrum L.*) también denominada *oro negro*, ya que, siglos atrás, había llegado a ser moneda de cambio entre romanos y godos [1]. Aunque no fue fácil encontrar a alguien dispuesto a sufragar los gastos de tal viaje, finalmente los Reyes Católicos aceptaron. Colón, en busca de una ruta alternativa al Oriente, se encontró con el Nuevo Mundo, y en lugar de la pimienta, encontró una especia proveniente del pimiento que denominaban *ají*, aunque él la denominó pimienta, generando así la confusión de estos vocablos que siguen presentes, incluso hoy día. No es de extrañar, por tanto, que se denomine al pimiento como la planta de los muchos nombres. Cristóbal Colón, a la vuelta de su viaje, dio a probar a los Reyes Católicos algunos de los frutos y productos que había traído de las Indias, entre ellos el pimiento. Francisco López de Gomara [2] cuenta cómo Isabel y Fernando probaron el nuevo fruto y les quemó la lengua. Muchos de los exploradores y cronistas que fueron llegando reportaron información sobre la forma de vida, las costumbres o la naturaleza del terreno americano. El Inca Garcilaso, en su crónica Comentarios Reales I [3] relata cómo los indios añaden ají a prácticamente todo lo que comen. José de Acosta dice en su libro Historia natural y moral de las Indias: “*Pero la natural especería que dio Dios a las Indias de occidente es la que en Castilla llaman pimienta de las Indias (...) [4].*

Hay indicios suficientes que revelan que la planta del pimiento es conocida desde tiempos muy remotos en América. Existen yacimientos arqueológicos que datan del 7000 a.C., donde se hallaron semillas y restos de frutos. Al igual que otras especies hortícolas, el pimiento se incorporó rápidamente a la lista de productos

cultivados en Europa, América Central, la India y Asia. En la figura I.1. se muestra el proceso de fabricación de uno de sus derivados principales: el pimentón.

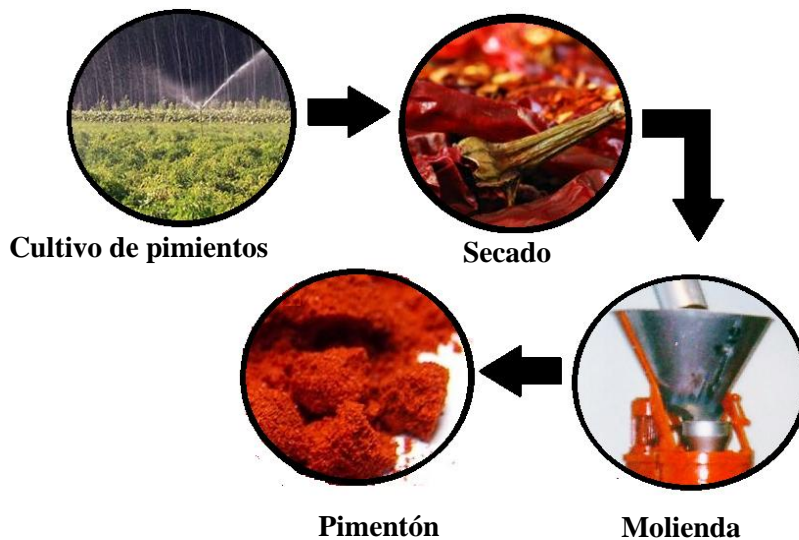


Figura I.1. Esquema de la producción de pimentón.

Se puede decir, que las especias y hierbas aromáticas han estado ligadas a la humanidad desde tiempos muy remotos para sazonar alimentos, conservarlos o como productos curativos.

A efectos del Real Decreto 2242/1984 [5], de 26 de septiembre, se designa con el nombre de especia o condimento aromático a las plantas o partes de las mismas, frescas o desecadas, enteras, troceadas o molidas, que por su color, aroma o sabor característicos se destinan a la preparación de alimentos y bebidas, con el fin de incorporarles estas características, haciéndolos más apetecibles y sabrosos y, en consecuencia, consiguiendo un mejor aprovechamiento de los mismos. Las especias provienen de las partes duras de las planta. Generalmente son semillas, como la mostaza, pero también pueden ser cortezas como la canela, o raíces en el caso del jengibre. Aunque aromatizan las preparaciones a las que se agregan, el uso de las especias influye más en el sabor que en la fragancia [6]. Las hierbas aromáticas, que son las hojas de los vegetales, como por ejemplo el orégano o la albahaca, tienen como función principal aromatizar, aunque también realzan el sabor de los alimentos [7].

I.1.2. Entrada del pimiento Europa

En la Europa del siglo XV, las especias tenían gran importancia, tanto por sus características de conservantes, como por influir directamente en los sabores de los alimentos [8]. La pimienta negra (*Piper nigrum* L.) era la más consumida, siendo un producto valioso de interés comercial entre Oriente y Occidente. La introducción del pimiento en Europa fue, en sus comienzos, una forma de sustituir a la pimienta y así acabó siendo. Entre los años 1569 y 1590 llegaban los barcos desde América a Europa, principalmente a Sevilla [9] cargados con pimientos que se distribuían por Europa, Asia y África.

El pimiento se difundió rápidamente en Europa. En el siglo XVI ya está documentada su presencia en países como Italia, Inglaterra o Hungría [10]. El Imperio Otomano tuvo un papel decisivo en la introducción y expansión del comercio de pimiento desde el golfo Pérsico hacia el Mediterráneo [11], por ello, no es de extrañar el amplio uso que se le da a este producto en Turquía o Grecia [1]. Hoy en día, en Europa los principales productores de pimiento son Grecia, Hungría, España, Rumanía, Italia y los Países Bajos.

I.2. Importancia del Género *Capsicum*

El género *Capsicum*, junto con otros 84 géneros más, constituye la familia Solanaceae, entre las cuales se encuentran el tomate o el tabaco. Se trata de una planta herbácea o arbusto. Dentro de este género existen 22 especies silvestres y 5 especies domesticadas que son *C. annuum* L., *C. baccatum* L., *C. chinense* Jacq., *C. frutescens* y *C. pubescens* que se cultivan en todo el mundo, siendo la *C. annuum* la más frecuente [12].

Dentro de estas especies existen subespecies que se clasifican por las características de los frutos tales como sabor picante [13], color, forma o tamaño. Al existir un gran número de subespecies, la terminología del género *Capsicum* se hace confusa. Para las plantas del género *Capsicum* podemos encontrar distintos nombres: pimientos, chilis, chile, ají o *Capsicum*, siendo, además, intercambiables [14].

El pimiento se considera una planta exigente en temperatura si se compara con el conjunto de hortalizas cultivadas en zonas similares. La temperatura es un factor influyente en el crecimiento, la fertilidad, e incluso en las dimensiones del fruto, hasta el punto de que éste no se desarrollará correctamente a menos que se den unas temperaturas determinadas. Si las temperaturas son demasiado bajas el fruto es delgado y puntiagudo, y si son demasiado altas el fruto es más redondeado [15]. También es muy sensible a los niveles de humedad relativa altos, siendo el nivel de humedad ideal del 70-75%. No presenta, sin embargo, exigencias especiales con la luz que recibe la planta [16].

Las especies del género *Capsicum* son muy consumidas en el mundo debido a su uso en diversos campos como la industria alimentaria [17], como colorante natural [18], en cosmética [19], en medicina debido a su actividad antitumoral [20], como componente en aerosoles de protección personal [21] o incluso como adsorbente para eliminar contaminantes [22]. La creciente demanda que este producto presenta se debe a varios factores entre los que cabe destacar el aumento de la población, la inmigración, sobre todo en países como los EE. UU y el cambio de preferencias alimentarias de los consumidores que apuestan más por los productos naturales frente a los artificiales.

Según la FAO [15], los principales países productores de *Capsicum* fresco en 2004 y 2011 fueron China, Turquía, México, Indonesia, España y EE.UU. En cuanto al pimiento seco, los países con mayor producción son India, China y Pakistán. En la figura I.2. se representan estos datos.

El comercio de productos de *Capsicum*, tanto fresco como seco, se recoge en las tablas I.1 y I.2. En el caso del pimiento fresco, los principales países importadores son EE.UU, Alemania y Reino Unido como se observa en la tabla I.1.

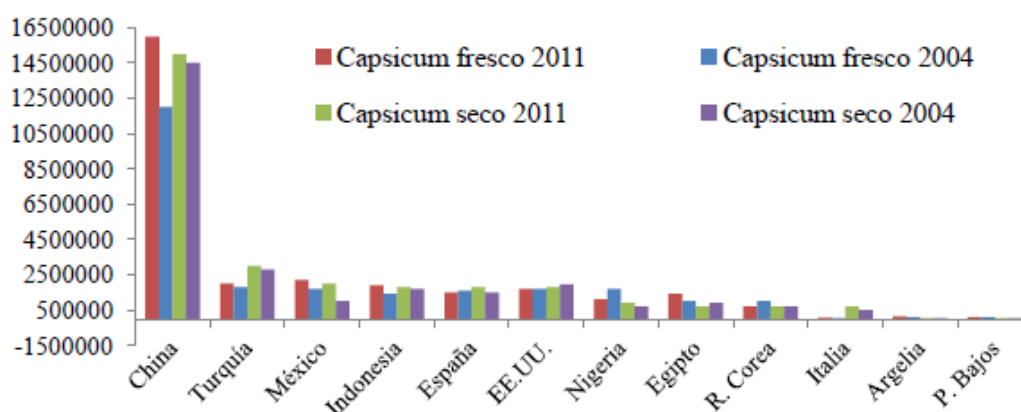


Figura I.2. Principales países productores de *Capsicum* fresco y seco.

Tabla I.1. Principales importadores de *Capsicum* fresco y seco

Países importadores			
Frutos secos		Frutos frescos	
Países	t	Países	T
EE.UU	109937	EE.UU	779393
Malasia	54296	Alemania	351622
Sri Lanka	42782	Reino Unido	157134
España	38141	Francia	136731
Tailandia	36970	Países Bajos	119896
México	24693	Canadá	118825
Alemania	20228	Rusia	114571
Indonesia	19988	Italia	96124
Rep. Corea	16309	Austria	51305
Bangladesh	13177	Rep. Checa	49548

t, tonelada métrica

Las exportaciones mayoritarias de pimiento fresco en 2011 corresponden a países como México, España o los Países Bajos. En cuanto al comercio del *Capsicum* seco, los principales importadores son EE.UU., Malasia y España. China, India y Perú, fueron los principales exportadores (tabla I.2.).

Tabla I.2. Principales exportadores de *Capsicum* fresco y seco

Países exportadores			
Frutos secos		Frutos frescos	
Países	t	Países	T
India	260485	México	699657
China	98479	España	511340
Perú	48471	Países Bajos	474013
España	34879	EE.UU	105379
Túnez	17451	Israel	101874
México	11007	Canadá	98113
Malasia	8817	China	85909
Alemania	5480	Turquía	68599
EE.UU	4802	Irán	64944
Países Bajos	3954	Marruecos	61539

t, tonelada métrica

En 2011 la superficie cultivada con especies del género *Capsicum* fue 1837704 Hectáreas. Los seis primeros países productores China, México, Turquía, Indonesia, EE.UU y España constituyen un 67% de esa superficie total.

Dado que en España la importancia del cultivo del pimiento es evidente, las principales zonas de cultivo y producción de derivados se unen al sistema de regulación de productos de calidad creado por la Unión Europea en 1992, para promover y proteger productos de una calidad determinada [23].

Hace algunos años se observó cómo los consumidores tendían a otorgar mayor importancia a la calidad de la alimentación. Esta búsqueda de productos específicos se reflejaba en una demanda de productos agrícolas y alimenticios de un origen geográfico determinado. Existía y existe una enorme variedad de productos comercializados, y gran cantidad de información sobre los mismos, ya que el consumidor debe poder disponer de datos claros y concisos acerca del origen del producto para poder elegir mejor. Asimismo, la producción, la fabricación y la distribución de los productos agrícolas y alimenticios ocupan un lugar importante en la economía de los países y, por tanto, es beneficiosa la promoción de los productos que presenten determinadas características para el mundo rural.

En este contexto las designaciones de Denominación de Origen Protegida (DOP) e Indicación Geográfica Protegida (IGP), son muy eficaces para garantizar la calidad y el origen geográfico de un producto. Según el Reglamento CEE (2081/92) [23] se entiende por:

- *Denominación de origen*: el nombre de una región, de un lugar determinado o, en casos excepcionales, de un país, que sirve para designar un producto agrícola o un producto alimenticio, que debe ser originario de dicha región, de dicho lugar determinado o de dicho país, y su calidad o características se deban fundamental o exclusivamente al medio geográfico con sus factores naturales y humanos, y cuya producción, transformación y elaboración se realicen en la zona geográfica delimitada.
- *Indicación Geográfica*: el nombre de una región, de un lugar determinado o, en casos excepcionales, de un país, que sirve para designar un producto agrícola o un producto alimenticio, que debe ser originario de dicha región, de dicho lugar determinado o de dicho país, y que posea una cualidad determinada, una reputación u otra característica que pueda atribuirse a dicho origen geográfico, y cuya producción y/o transformación y/o elaboración se realicen en la zona geográfica delimitada.

Para acogerse a una DOP o una IGP, un producto agrícola o alimenticio deberá ajustarse a una serie de condiciones que se proponen y se corrigen en el Reglamento (CE) Número 510/2006 [24].

Las DOP relacionadas con el pimentón en España son las siguientes:

- DOP. <<Pimentón de la Vera>> (Extremadura) [25]: En el Reglamento (CE) No 982/2007 de 21 de agosto de 2007 [26] se inscribe el<<pimentón de la Vera>> como DOP. Los pimientos usados para elaborar pimentón de la Vera, deben pertenecer a la zona delimitada por las cuatro comarcas naturales del norte de la provincia de Cáceres: La Vera, Campo Arañuelo, Valle del Ambroz y Valle del Alagón. En el apartado I.3.2.1 se desarrollará en detalle esta DOP.

- DOP. <<Pimentón de Murcia>> (Murcia) [27]: La zona geográfica de producción la constituyen 14 municipios que son: Alhama de Murcia, Beniel, Cartagena, Fortuna, Fuente Álamo, Librilla, Lorca, Molina de Segura, Murcia, Puerto Lumbreras, San Javier, Santomera, Torre Pacheco y Totana, de la provincia de Murcia. En el apartado I.3.2.2 se desarrollará en detalle esta DOP.

En España existen, además, otras DOP e IGP relacionadas con el pimiento

- DOP. <<Pimiento del Piquillo de Lodosa>> (Navarra) [28].
- IGP. <<Pimiento Riojano>> (La Rioja) [29, 30].
- IGP.<<Pimiento Asado del Bierzo>> (Castilla y León) [31]
- IGP.<<Pimiento de Fresno-Benavente>> (Castilla y León) [32].
- IGP.<<Pimiento de Guernica" / "Gernikako piperra>> (País Vasco) [33]

En la figura I.3 se muestran las etiquetas correspondientes a las DOP de pimentón presentes en España. El etiquetado de este tipo de productos está regulado y es uno de los parámetros de control.



Figura I.3. Etiquetas para las DOP de pimentón: La Vera y Murcia.

I.3. Pimentón. Definición, elaboración y tipos

I.3.1. Definición y elaboración

Se considera pimentón a aquella especia obtenida tras el secado y molienda de los frutos del género *Capsicum* que se encuentra dentro de la familia de las Solanáceas. El secado puede efectuarse según diversas técnicas, obteniendo

resultados diferentes. En cierto momento de la producción es posible la adición de aceites vegetales en las proporciones previamente establecidas y legisladas. La adición de estos aceites no influye en el carácter del pimentón [25, 27].

I.3.2. Denominaciones de origen de pimentón en España

Las primeras noticias sobre este cultivo en la provincia de Cáceres datan de finales del siglo XV. Fue iniciado por los frailes Jerónimos del Monasterio de Guadalupe; así llegó hasta el Monasterio de Yuste (Cuacos de Yuste), desde donde se trasladó a Murcia a través del convento de esta Orden en la Ñora [34].

Así se definieron las dos principales zonas de producción de pimentón de España: La Vera en Cáceres, (Extremadura) y Murcia. A pesar de tener un origen común y un desarrollo prácticamente paralelo, el proceso de producción es diferente en cada una de estas zonas. En ambos casos el producto es resultado del secado y molienda de frutos del género *Capsicum*, pero cambia el tipo de secado y de variedad de fruto, haciéndolo así también las características organolépticas.

I.3.2.1. Pimentón de La Vera (Extremadura)

Las noticias sobre el cultivo de pimiento en la Vera son bastante antiguas. En los Interrogatorios de la Real Audiencia de Extremadura, en 1753, citan al pimiento entre los cultivos principales, en Jaraíz de la Vera.

El proceso industrial para la obtención de pimentón se inicia a finales del siglo XVII, alcanzando un importante desarrollo a mediados del siglo XVIII. Los pimientos, entonces, se trituraban en los molinos hidráulicos de harina que se encontraban en los márgenes de las gargantas. El cultivo se asienta definitivamente a mediados del siglo XVIII en pequeños bancales en el margen derecho del río Tiétar. La expansión continuó durante el siglo XIX, hasta convertirse en el cultivo principal de la comarca a finales de siglo, desplazando a otros productos tradicionales en La Vera como el lino o la seda [35]. La llegada de la luz eléctrica a La Vera permitió el

uso de molinos eléctricos y, como consecuencia, una sustancial mejora en las industrias, que comienzan a instalarse con el objetivo único de elaborar pimentón.

En el año 1994 se aprueba con carácter provisional la notación de DOP «Pimentón de la Vera», haciéndose definitiva tras la aprobación de su Reglamento el 21 de mayo de 1998. Finalmente, la DOP «Pimentón de la Vera», cuyo Reglamento se aprueba por el Ministerio de Agricultura, Pesca y Alimentación mediante la Orden APA 4/4.178/2005 de 22 de Diciembre [25] y por la UE [26] en diciembre de 2007, suponiendo la protección y reconocimiento de este producto en el marco de la UE.

Según el citado Reglamento de la Denominación de Origen Protegida «Pimentón de la Vera» y su Consejo Regulador, quedan protegidos con la Denominación de Origen Protegida «Pimentón de la Vera» los pimentones tradicionalmente designados bajo esta denominación geográfica que, reuniendo las características especificadas en este Reglamento, hayan cumplido en su producción, secado y elaboración todos los requisitos exigidos en el mismo y en la legislación vigente.

La zona de producción de pimientos para la elaboración del «Pimentón de la Vera» está constituida por cuarenta y nueve municipios pertenecientes a cuatro comarcas naturales en el norte de la provincia de Cáceres: La Vera, Campo Arañuelo, Valle del Ambroz y Valle del Alagón. En la figura I.4. se muestra un mapa de España con las dos zonas correspondientes a las DOP, Extremadura (Pimentón de La Vera) y Murcia (Pimentón de Murcia), destacadas en verde y amarillo, respectivamente.

Los frutos rojos empleados para la producción de «Pimentón de La Vera» pertenecen a las variedades del grupo de las Ocales (Jaranda, Jariza y Jeromín) y Bola, dentro de las especies *Capsicum annuum L.* y *Capsicum longum L.*, respectivamente. Dependiendo el tipo de pimentón: dulce, picante o agridulce (también conocido como ocal), se utilizan distintos frutos. Así, para el tipo dulce se utilizan los pimientos de la variedad Jaranda y Bola, para el tipo picante, Jeromín, Jariza y Jaranda, y para los agridulces u ocales, Jaranda y Jariza.



Figura I.4. Mapa de España con las principales zonas de producción de pimentón.
En verde Cáceres (Extremadura), en amarillo Murcia.

Otra de las características del pimentón producido en La Vera es el procedimiento de secado al humo. Se considera la primera fase de elaboración del pimentón y se lleva a cabo en secaderos de corriente vertical con hogar inferior, donde los frutos se secan con los gases de la combustión lenta de madera de encina y roble. Este sistema se desarrolla ante la imposibilidad de secado al sol, debido a las lluvias otoñales que coinciden con el tiempo de cosecha. Este secado artesanal, lento y suave, le confiere al pimentón de la Vera un aroma y sabor característicos.

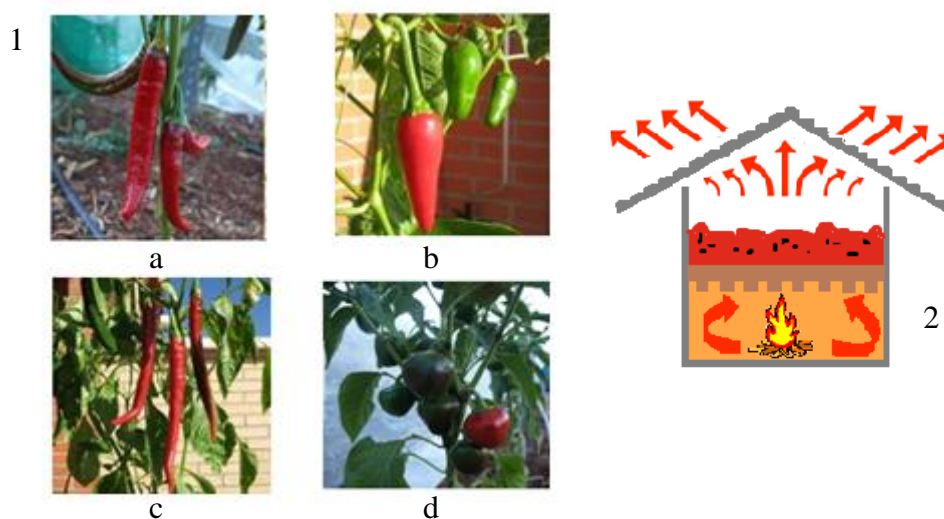


Figura I.5. 1. Tipos de frutos. a: Jaranda; b: Jeromín; c: Jariza; d: Bola. 2. Horno de corriente vertical con hogar inferior.

En la figura I.5. se muestran distintas variedades de pimientos así como un esquema del secado propio de la DOP de La Vera.

I.3.2.2. Pimentón de Murcia (Murcia)

Las primeras noticias que se tienen sobre el cultivo del pimiento en la Región de Murcia se sitúan a comienzos del siglo XVI, y son los frailes Jerónimos los responsables de su introducción, al cultivarlos en unas parcelas que poseían cerca de la zona. Las semillas traídas de América proceden de un pimiento ligeramente alargado y picante, de carácter silvestre y, por ello, algunos autores comentan que las características originales de este tipo de pimiento han sido influidas por las condiciones medioambientales de Murcia y las técnicas culturales de cultivo, dando lugar a que en unos años se suavizara su forma y sabor, siendo más redondeado y algo menos picante [36]. A mediados del siglo XIX fue cuando se inició la industrialización del Pimentón en la Región de Murcia. Durante el XIX y mediados del XX, el cultivo de pimientos para pimentón se fue extendiendo por toda la Vega del Segura, llegando a ser la actividad económica regional más importante. En 1986 se constituye la primera asociación del gremio del pimentón bajo el nombre “Centro de Exportadores de Pimiento”

El 1998 se solicitó la protección del pimentón de Murcia y finalmente el 17 de diciembre de 2001 se creó el Consejo Regulador de la Denominación de Origen Protegida. DOP «Pimentón de Murcia» [27].

La zona geográfica de producción está constituida por los terrenos ubicados en Alhama de Murcia, Beniel, Cartagena, Fortuna, Fuente Álamo, Librilla, Lorca, Molina de Segura, Murcia, Puerto Lumbreras, San Javier, Santomera, Torre Pacheco y Totana, de la provincia de Murcia.

Los frutos empleados para la producción del pimentón con denominación de origen Murcia son frutos rojos del género *Capsicum annum* L., variedad Bola. Son pimientos dulces y de poco peso.

En Murcia, como el clima es seco y cálido, es posible el tradicional secado de pimientos al sol, lo que le proporciona un color rojo brillante característico frente a otros obtenidos en secaderos de aire caliente [37,38]. La técnica de secado al sol consiste en extender los pimientos enteros sobre eras y dejarlos expuestos al sol de 1 a 4 días. Existe también un sector agricultor en Murcia que realiza el secado en secaderos de aire caliente con las fases previas correspondientes de lavado y cortado. La normativa impide superar la temperatura crítica de 65-75 °C en esta operación, lo que supone tiempos de 8-10 h de secado. A la salida del producto del secado, la temperatura del aire debería estar alrededor de 40 °C [39].

I.4. Parámetros de calidad del pimentón

En los respectivos documentos oficiales [25,27] para la denominación de origen, elaboración, circulación y comercio de Pimentón de la Vera y Pimentón de Murcia, vienen recogidas las cantidades toleradas para algunos elementos, así como los parámetros de calidad considerados. Estos parámetros se recogen en la tabla I.3.

Tabla I.3. Parámetros de calidad en pimentones.

Parámetro	La Vera (Extremadura)	Murcia
Dimensión de grano	16ASTM*	16ASTM*
Humedad máxima	14%	14%
Fibra bruta sobre materia seca	0,28	< 28%
Extracto etéreo sobre materia seca	0,23	< 23%
Cenizas sobre materia seca total	0,01	< 0,7%
Color ASTA mínimo	90	> 90

*equivalente a 1,19 mm de luz de malla.

No deben contener un residuo de metales superior a 1 ppm para el arsénico y 4 ppm para el plomo. En cuanto a los ingredientes alimentarios añadidos, se permite la adición de aceites vegetales comestibles en una proporción máxima del 8% sobre el producto seco.

I.5. Composición mineral y orgánica

Las propiedades organolépticas de los productos que consumimos, cada vez cuentan con una mayor atención por parte de los consumidores y los productores [40]. Las propiedades organolépticas de las muestras vegetales dependen en gran medida de su composición [41], que varía en función de la especie, de las condiciones de crecimiento, el estado de maduración, etc...

El agua es el componente más abundante en muestras vegetales, siendo en ocasiones el 90% de la masa total. Además del agua existen dos grandes grupos a tener en cuenta en el estudio de la composición de muestras vegetales: componentes orgánicos e inorgánicos. Atendiendo a la fracción orgánica, cabe destacar la presencia de ácidos orgánicos, lípidos, ácidos grasos, carbohidratos, fibra, vitaminas, capsaicinoides y compuestos volátiles.

La presencia de ácidos orgánicos varía en función de la especie. Algunos de los principales son el decanoico y dodecanoico, y algunos aromáticos como el quínico y el benzoico [42]. La concentración de los ácidos desciende con el proceso de maduración. La presencia de ácidos grasos y lípidos [43] varía según el fruto del género *Capsicum* que se estudie. En muestras vegetales, en general, la concentración de ácidos grasos es muy baja, a excepción de aguacates o aceitunas [44]. Además de ácidos orgánicos existen otros compuestos orgánicos volátiles que son los responsables del aroma del pimiento. Existen ciertas diferencias en la presencia y concentración de éstos en el pimiento fresco y el procesado, casi siempre por motivos del secado del fruto. En la fracción volátil la concentración de pirazinas no es muy elevada; sin embargo, el 2-isobutil-3-metoxipirazina posee un fuerte aroma característico a pimiento, que incluso tras el proceso de secado es detectable [45]. Los terpenos y sus derivados, entre los que están los carotenoides, están presentes en los frutos frescos en concentraciones más altas que en los procesados, aunque también juegan un papel importante en la composición del pimiento procesado, como lo es el pimentón. Existen también un gran número de alcoholes, aldehídos, cetonas y ésteres entre otros. En total hay 125 compuestos orgánicos volátiles identificados presentes en el pimiento [46]. Los carbohidratos son los componentes orgánicos mayoritarios de las muestras vegetales, y por tanto del pimiento. Las funciones de los

carbohidratos incluyen, entre otras, el almacenaje de la energía, y son responsables de algunos atributos sensoriales. La fibra dietética que se define como la cantidad de carbohidratos no digeribles y la lignina intrínseca e intacta en plantas [47] está presente en el pimiento con cantidades aproximadas del 2,1% [48].

Otro conjunto importante de moléculas orgánicas son las vitaminas, necesarias para el desarrollo pero que, generalmente, no se sintetizan de forma natural, sino que hay que aportarlas al organismo a través de los alimentos. El pimiento contiene una serie de pigmentos denominados carotenoides. A los que contienen átomos de oxígeno se les conoce más específicamente como xantofilas y al resto se les denomina carotenos. Los principales son β -caroteno, luteína, capsantina y capsorrubina, que son los encargados de activar la vitamina A, que juega un papel importante en la visión, división celular y el desarrollo de los huesos. En general, se acumulan más en la piel que en el fruto en sí [49]. Existen cambios en la presencia de estos carotenoides en el proceso de recolecta y tratamiento del pimiento [50]. Phillips et al. [51] estudiaron la vitamina B9 (ácido fólico), importante en la dieta, y concluyeron tras realizar análisis por HPLC que el pimiento rojo tiene una concentración de folatos totales mayor que el pimiento verde. Daood et al. [52] investigaron el efecto que tiene el secado de pimientos en la vitamina C, entre otras, y concluyeron que la retención de ascórbico en pimientos secados de forma natural era más baja que cuando se secaron con corrientes de aire. La vitamina E, también importante como antioxidante, incluye los tocoferoles y tocotrienoles, existiendo diversos isómeros, siendo el más efectivo como antioxidante el α -tocoferol [53]. La concentración de tocoferoles se encuentra en el rango de 3,7 a 236 mg por 100 gramos de fruto seco [54] y al igual que la vitamina C son inhibidores de la oxidación lipídica, que, entre otros efectos, causa la pérdida de color de los frutos [55]. Los pimientos, especialmente los rojos, son una fuente rica en vitamina A, C y E, por lo que se consideran alimentos con buenas propiedades antioxidantes.

Los capsaicinoides son los compuestos responsables del sabor picante de los pimientos [56]. Se trata de un grupo de amidas ácidas formadas a partir de vanillilamina y ácidos grasos de 8 a 13 átomos de carbono. La capsaicina y la dihidrocapsaicina constituyen aproximadamente el 90% de los capsaicinoides totales

presentes en las variedades picantes. Se han encontrado más de 13 compuestos capsaicinoides, siendo la nordihidrocapsaicina el tercero más abundante junto con los anteriormente nombrados. La primera vez que se determinó el sabor picante, fue por un test organoléptico denominado escala Scoville (Scoville Heat Units, SHU) [57]. La tendencia en determinación del grado de picor ha sido el análisis por técnicas analíticas como HPLC o GC.

Los capsinoides, conocidos también como Capsaicinoid-like substances (CLS) por su parecido con los capsaicinoides, son compuestos naturales que se encuentran en algunas variedades de pimientos dulces. Son ésteres formados a partir de la condensación del alcohol vainílico y ácidos grasos. La diferencia estructural con respecto a los capsaicinoides es el enlace del esqueleto bencénico con la cadena carbonada, ya que en los capsinoides es tipo éster mientras que en los capsaicinoides es tipo amida. En cuanto a propiedades organolépticas, los capsinoides no presentan sabor picante. En la figura I.6. se muestra la estructura de los principales capsaicinoides y capsinoides presentes en el pimiento.

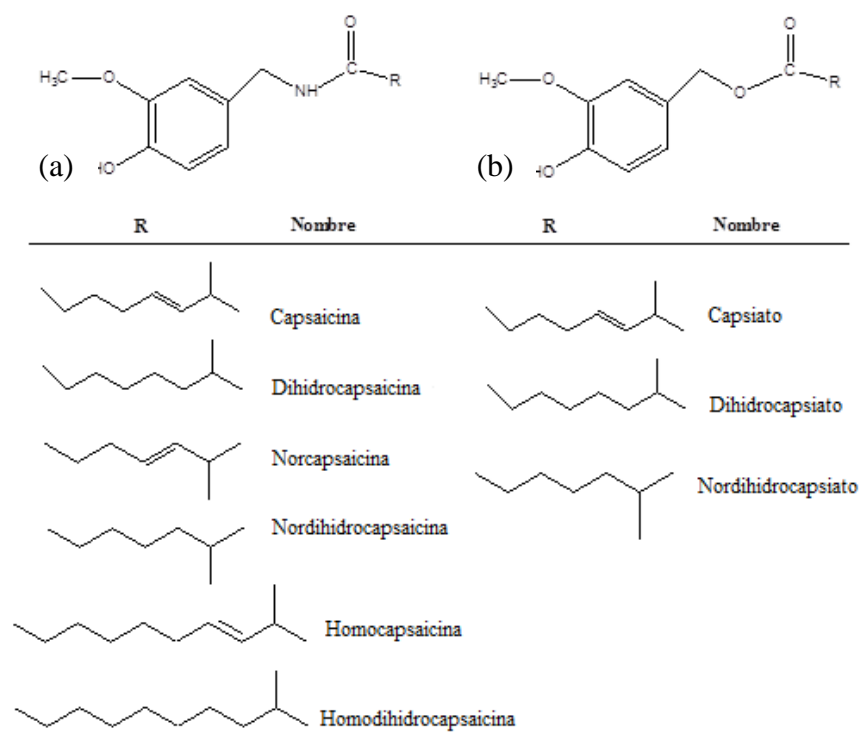


Figura I.6. Capsaicinoides (a) y capsinoides (b) más comunes en el pimentón.

En cuanto a la fracción inorgánica en muestras vegetales, se puede hablar de la presencia de elementos, también denominados macronutrientes, que son aquellos presentes en mayor concentración. Éstos son Ca, K, Mg, N y P. Los micronutrientes están presentes en concentraciones más bajas y son Cl, F, Fe, I, Mn, Na y S. También, hay elementos traza presentes a muy bajas concentraciones como Co, Cu, Se y Zn.

La presencia de metales en muestras vegetales tiene distintos orígenes. Por lo general, se debe a factores naturales y a factores antropogénicos. Los factores naturales se relacionan con la genética de la planta, el clima y la composición del suelo. Witney et al. [58] explicaron cómo se distribuyen los elementos Ca, K y Mg en plantas con distintas características. Tukey [59] explicó cómo las plantas absorben minerales del suelo mediante soluciones acuosas como el agua de lluvia o la niebla. Por otro lado, los factores antropogénicos o exógenos se relacionan con el uso de fertilizantes, insecticidas, funguicidas, procesos de secado y otras fuentes de contaminación [60]. Estrada et al. [61] estudiaron el efecto que pueden tener algunos nutrientes inorgánicos presentes en los fertilizantes en el picante de las muestras.

1.5.1. Métodos de análisis de compuestos orgánicos en pimiento y pimentón.

La fracción orgánica es la responsable de gran parte del olor y el sabor de las muestras vegetales, así como de ciertas propiedades nutricionales, dietéticas e incluso médicas. Por ello, el estudio de los componentes de las muestras, en este caso, de pimiento y pimentón es de gran importancia. Las principales sustancias que se determinan son los carotenoides, compuestos volátiles, capsaicinoides y capsinoides.

La preparación previa de la muestra es un requisito para cualquier proceso de análisis y consiste, por lo general, en extracción, pre-concentración y filtrado de las muestras. La extracción de capsaicinoides y capsinoides en pimientos y pimentones se ha llevado a cabo con diferentes disolventes y métodos de extracción. Los disolventes más empleados son orgánicos como metanol [62], etanol [63], acetonitrilo [64], acetona [65], acetato de etilo [66], cloroformo [67] o mezcla de varios disolventes como acetato de etilo y éter etílico [68]. Se han utilizado diversos

métodos de extracción de capsaicinoides, como la maceración [64], extracción asistida por ultrasonido [69], extracción mediante agitación magnética [71] o extracción en Soxhlet [65] y extracción enzimática [72], con fluidos supercríticos o asistida por microondas [73]. Para los carotenoides las extracciones, son muy parecidas. Breithaupt y Schwack [74] determinaron el contenido de carotenoides por cromatografía de gases y cromatografía líquida, previa extracción y transesterificación de la muestra usando para la extracción una mezcla de metanol, acetato de etilo y éter de petróleo ligero, evaporando el disolvente y redisolviendo el extracto en metanol y *tert*-butil éter. El extracto se somete a transesterificación para la correcta identificación de los carotenoides. Cámara et al. [75] extrajeron los carotenoides de las muestras mediante el uso de distintos disolventes orgánicos separando los carotenos de las xantofilas y clorofilas en una columna de celulosa.

1.5.1.1 Determinaciones mediante cromatografía líquida de alta resolución

La cromatografía líquida de alta resolución, HPLC, es una de las técnicas más utilizadas en la determinación de compuestos orgánicos en muestras vegetales y de especias. Su campo de aplicación cubre una gran parte de compuestos, entre ellos aquellos que son termosensibles, con masas moleculares grandes o compuestos polares. El principal objetivo relacionado con esta técnica en el análisis de pimientos y derivados es la determinación y cuantificación de capsaicinoides y capsinoides [76].

Los trabajos publicados sobre de este tema sugieren prácticamente en su totalidad el uso de fases estacionarias apolares, cromatografía en fase inversa, para la separación y cuantificación de las especies capsaicinoides y capsinoides. No obstante, algunos autores como, Iwai et al. [66] desarrollaron un método para la determinación de capsaicinoides en muestras de pimientos, trabajando en fase normal y régimen isocrático con fase móvil mezcla de isopropanol:n-hexano:metanol. Mediante elución en fase inversa Collins et al. [77], desarrollaron métodos para determinar y cuantificar los cuatro capsaicinoides principales: capsaicina, dihidrocapsaicina, nordihidrocapsaicina y homodihidrocapsaicina en pimientos, empleando como fase móvil una mezcla metanol-agua (73:27 v/v). Al Othman et al.

[78] determinaron el contenido en capsaicina y dihidrocapsaicina en pimientos mediante HPLC con detección UV en fase inversa utilizando como fase móvil una mezcla de acetonitrilo: agua en régimen isocrático y el grado de picor de las muestras [79]. Los límites de detección para el método propuesto fueron de 0,09 y 0,1 $\mu\text{g g}^{-1}$ para la capsaicina y dihidrocapsaicina, respectivamente. Los límites de cuantificación para la capsaicina y dihidrocapsaicina fueron 0,3 y 0,36 $\mu\text{g g}^{-1}$.

Perucka et al. [68] propusieron un método de determinación de capsainoides en frutos de pimiento mediante HPLC en fase inversa usando un gradiente de fase móvil acetonitrilo: agua. La extracción y separación previa de los capsainoides y carotenoides, se lleva a cabo mediante cromatografía en capa fina [80]. Schweiggert et al. [81] desarrollaron un método para separar y cuantificar mediante elución en gradiente los capsainoides mayoritarios y minoritarios usando como fase móvil un disolvente A consistente en 0,5% de ácido acético en agua: acetonitrilo (90:10 v/v) y uno B mezcla de un 0,5% de ácido acético en agua: acetonitrilo (10:90 v/v).

Juangsamoot et al. [71] validaron un método para la determinación del contenido de capsaicina y dihidrocapsaicina en algunas variedades de pimientos mediante HPLC en fase inversa con detector de fila de diodos. Obtuvieron unos límites de detección y cuantificación de 0,5 y 1 $\mu\text{g g}^{-1}$, respectivamente, y unos valores de repetitividad y reproducibilidad aceptables. En este caso, la extracción previa se hizo mediante agitación magnética y posterior limpieza en cartuchos de extracción en fase sólida. Betts [82] propuso una extracción en fase sólida con C18 para la extracción de los capsainoides de salsas picantes a base de pimiento. La determinación se realiza después mediante HPLC con detector UV a 280 nm.

Peusch et al. [83] desarrollaron un método para la determinación de los capsainoides en pimentones de bajo sabor picante que implicaba una elaborada preparación de la muestra. En primer lugar se estudió una primera purificación con óxido de aluminio [84] para la oleoresina de pimentón, mientras que para el extracto se propone un tratamiento con gel de sílice.

Garcés-Claver et al. [85] desarrollaron y validaron un método analítico para identificar y cuantificar de manera adecuada algunos capsainoides mediante HPLC

acoplado a un espectrómetro de masas con ionización por electrospray y detector de tiempo de vuelo (HPLC-ESI-MS/TOF), consiguiendo límites de detección mejores que los obtenidos al utilizar HPLC acoplado a un detector UV-Vis. La resolución del MS/TOF proporciona información de la distribución isotópica, lo que se puede emplear para estudios metabólicos, por ejemplo en plantas sometidas a diferentes tipos de cultivos. Mendel et al. [86] validaron un método de HPLC para el análisis de ocho capsaicinoides en diferentes partes del pimiento, modificando los porcentajes de composición de fase móvil que influyen en los tiempos de retención y los factores de capacidad. La identificación de los picos se llevó a cabo usando MS, y la cuantificación por medio de la respuesta obtenida en el HPLC usando un detector UV a 280 nm.

Singh et al. [87] propusieron un método para la determinación de los capsinoides mediante HPLC con detector de fila de diodos, en régimen isocrático y fase móvil compuesta por acetonitrilo-agua (60:40) con un flujo de 1 ml min^{-1} , utilizando una columna monolítica C-18 para la correcta separación de dichos capsinoides y aprovechando las ventajas que proporcionan en cuanto al tiempo de análisis [88].

Los carotenoides se han determinado en diversas muestras de alimentos. Hart y Scott [89] desarrollaron un método para el análisis de carotenoides en alimentos y muestras vegetales, entre ellas el pimiento, y frutas consumidas en el Reino Unido. Tras una fase de extracción y de saponificación se determinaron los carotenoides: luteína, zeaxantina, β -criptoxantina, licopeno, β -caroteno y α -caroteno mediante HPLC con una fase móvil consistente en acetonitrilo, metanol, diclorometano (75:20:5). Topuz et al. [90] determinaron el contenido de carotenoides, también mediante HPLC, con el fin de estudiar la influencia que ejerce el tipo de secado de pimiento, en los carotenoides y capsaicinoides. Las muestras se trataron previamente y se saponificaron antes de proceder a la determinación en HPLC. Se trabajó en fase inversa en régimen de gradiente con una fase móvil compuesta de acetona-agua en el inicio y una velocidad de flujo de 1 mL min^{-1} . Determinaron los compuestos a 450 nm, y concluyeron que el tipo de secado influía notablemente en el color de la muestra. Mínguez-Mosquera et al. [91] separaron y cuantificaron varios carotenoides

en muestras de pimientos del tipo Bola y Agridulce. Se hicieron medidas tanto de extractos saponificados como directos. Existe cierta controversia en el tema, pues algunos autores afirman que la saponificación es un procedimiento agresivo que altera la muestra [92], mientras que otros como Mózsik et al. [93] mantienen que si la saponificación es controlada, no provoca transformaciones apreciables en la misma.

Recientemente, Giuffrida et al. [94] han caracterizado 12 variedades del género *Capsicum* y han determinado 52 carotenoides distintos mediante HPLC con un detector de fila de diodos acoplado a un espectrómetro de masas con ionización química a presión atmosférica.

1.5.1.2. Determinación mediante cromatografía de gases.

La cromatografía de gases, GC, también se ha utilizado para la determinación de capsaicinoides. El análisis está limitado por la baja volatilidad de estos compuestos y es necesario derivatizarlos antes de la inyección.

Muller-Stock et al. [95] desarrollaron un método para determinar cuantitativamente la capsaicina y otros derivados realizando una trimetilsililación de los compuestos. Karajewska y Powers [96] determinaron y cuantificaron capsaicinoides presentes en muestras de pimientos mediante pentafluorobenzilación y determinación con un detector de captura electrónica, CE, obteniendo límites de detección inferiores a 10 pg mL^{-1} para la capsaicina.

Thomas et al. [97] desarrollaron un método de GC sin derivatización, usando una columna polar y con un detector termoiónico selectivo.

La GC se emplea comúnmente en la determinación de los compuestos volátiles presentes en las muestras y responsables en gran medida del aroma [98]. Guadayol et al. [99] estudiaron cinco procedimientos de extracción de la fracción volátil de muestras de oleorresina de pimentón español: espacio de cabeza dinámico y estático, purga y trampa y destilación-extracción simultánea a presión reducida y a presión atmosférica, siendo estos últimos los más convenientes para evitar degradación térmica. Identificaron los compuestos volátiles mediante GC-MS. Para probar la eficiencia de las extracciones se recurrió a la HRGC-FID. Oruña-Concha et

al. [100] realizaron la extracción de volátiles en pimientos de padrón mediante espacio de cabeza dinámico, y analizaron la fracción volátil mediante GC-MS.

Pino et al. [101] identificaron 125 compuestos volátiles de pimientos del tipo Habanero en dos fases distintas del cultivo, usando un FID como detector y confirmación mediante espectrometría de masas. Estos mismos autores [102] caracterizaron pimientos Habaneros cultivados en Yucatan en términos de compuestos volátiles, que se obtuvieron por destilación-extracción en corriente de vapor, y se analizaron mediante GC-FID y GC-MS.

Buttery et al. [103] aislaron la fracción volátil mediante un sistema de extracción continua en corriente de vapor a vacío utilizando hexano como disolvente, y en corriente de vapor a presión atmosférica usando pentano. Caracterizaron las muestras de pimientos verdes mediante GC-MS, interpretando los espectros de masas y comparándolos con una muestra ya caracterizada previamente en el mismo equipo. Además, utilizaron técnicas como espectroscopía infrarroja (IR) y resonancia magnética nuclear (RMN) para confirmar las sustancias. Los compuestos más abundantes encontrados fueron, limoneno, trans- β -ocimeno, metil salicilato y 2-metoxi-3-isobutilpirazina.

Kocsis et al. [104] investigaron el aroma de pimentones de origen húngaro comparando el perfil de compuestos volátiles de un pimentón picante y un dulce, extracción se llevó a cabo mediante destilación en corriente de vapor y la identificación mediante GC-MS.

Kang y Baek [105], analizaron el perfil aromático de una salsa a base de pimiento rojo, mediante SPME del espacio de cabeza y GC acoplada a una nariz electrónica.

Ziino et al. [106] determinaron el contenido de la fracción volátil de muestras del género *Capsicum annuum* L., pertenecientes a la zona de Calabria, mediante GC-MS. La extracción de los compuestos volátiles se llevó a cabo mediante SPME espacio de cabeza.

I.5.2. Determinación de metales en pimiento y pimentón

Las técnicas más frecuentes en la determinación de metales en muestras vegetales son la espectroscopía de absorción y la de emisión atómica. Un paso crucial antes de llevar a cabo estos análisis es la preparación de la muestra, que generalmente requiere una completa digestión o mineralización [107]. La mineralización se puede realizar: por vía húmeda o vía seca siendo éste último el menos aconsejado por algunos autores [108, 109]. Dentro de los métodos de digestión por vía húmeda encontramos que pueden realizarse en sistemas abiertos [110], o sistemas cerrados [111], y en ambos casos la presión puede variar y pueden ser digestiones asistidas por microondas o ultrasonido. Aydın [112] comparó los métodos de digestión por vía seca, por vía húmeda y en microondas. Altundag y Tuzen [113] también compararon los efectos de realizar las digestiones de muestras vegetales por vía seca, húmeda o asistida por microondas. Filgueiras et al. [114] compararon los resultados de extracciones de muestras vegetales usando digestión asistida por microondas frente a una digestión asistida por ultrasonido. Los resultados llevaron a concluir que la digestión asistida por ultrasonido presentaba ventajas frente a la extracción vía microondas, como el menor tiempo de tratamiento, menor cantidad de reactivos usados y mayor seguridad en cuanto a la operación. Lavilla et al. [115] desarrollaron distintos métodos de digestión a baja, media y alta presión asistida por microondas y con calentamiento en baño de arena.

Antonious et al. [116] determinaron el contenido de Cu, Cd, Cr, Mo, Ni, Pb y Zn mediante espectroscopía de emisión atómica de plasma inducido acoplado (ICP-OES) para evaluar el transporte de dichos metales desde la tierra al fruto del pimiento, pues dependiendo del tipo de variedad de pimiento se acumula en mayor proporción un metal u otro [117]. Khairiah et al. [118] determinaron el contenido de Cd, Pb y Zn en muestras vegetales, entre ellas de pimientos, para estudiar la especiación y toxicidad de dichos metales pesados en zonas agrícolas y el transporte de la tierra al fruto, mediante espectroscopía de absorción atómica (AAS), previa digestión por vía húmeda con mezcla de ácidos oxidante.

Divrikli et al. [119] determinaron el contenido de Cd, Cr, Cu, Fe, Mn, Ni, Pb y Zn en especias y hierbas naturales mediante espectroscopía de absorción atómica

de llama (FAAS), previa digestión ácida, y estudiaron la influencia de dichos metales en las muestras. Aydin [112] determinó el contenido de Cd, Co, Cr, Cu, Fe, Mg, Mn, Ni, Pb y Zn mediante ICP-OES y empleó FAAS para la determinación de Na, Ca y K. Las muestras de origen natural requerían pre-tratamiento y se compararon distintos métodos de digestión. Soylak et al. [120] evaluaron el contenido de Co, Cr, Cu, Fe Mn, Ni, Pb y Zn en siete tipos de plantas aromáticas y especias cultivadas en Turquía, entre ellas *Capsicum annuum* L., mediante FAAS después una digestión ácida por vía húmeda.

Díaz-Alarcón et al. [121] determinaron el contenido de Se en muestras de especies diferentes, entre las que se encuentra el pimiento, *Capsicum annuum* L. del sudeste de España mediante espectroscopia de absorción atómica con generación de hidruros (HG-AAS). El valor medio de la concentración encontrada para el pimiento es de 2,12 ng g⁻¹ de producto fresco.

Sziklai et al. [122] determinaron la distribución de algunos micronutrientes y metales trazas en plantas del genero *Capsicum annuum* procedentes de Hungría mediante espectroscopía de rayos γ usando un detector de germanio conectado a un analizador multicanal. Chizzola et al. [123] estudiaron los micronutrientes y metales pesados de hierbas, plantas medicinales y especias consumidas con asiduidad en Austria, mediante AAS. Cu, Mn, Fe y Zn se analizaron mediante la técnica de llama, y Cd y Pb con atomización electrotérmica.

1.5.3. Otras técnicas usadas en el análisis de pimiento y pimentón

La electroforesis capilar es una técnica de separación muy usada para la determinación de la composición orgánica y mineral de diversos tipos de muestras. Para muestras del género *Capsicum*, Liu et al. [124] separaron, identificaron y cuantificaron capsaicina y dihidrocapsaicina en muestras del género *Capsicum*, mediante electroforesis capilar (CE) y una mezcla de surfactantes, en salsas donde el pimiento es el ingrediente principal y en muestras de *C. annuum*. Laskaridou-Monnerville [125] también utilizó CE para la determinación de capsaicina y dihidrocapsaicina en varias especies del género *Capsicum*. A diferencia de Liu et al.,

que tras la extracción únicamente microfiltraban la muestra, Laskaridou-Monnerville pasaba la muestras por una columna de C-18 antes del análisis por CE.

Bajaj y Kaup [126] desarrollaron un método colorimétrico para la determinación de capsaicina en frutos del género *Capsicum* consistente en una extracción, una purificación en columna de alúmina y finalmente la medida del color desarrollado mediante el reactivo de Folin-Ciocalteu, midiendo la absorbancia a 760 nm. Gibbs y O'Garro [127] propusieron modificaciones al método desarrollado por Bajaj y Kaup [126] que permitía la supresión de la etapa de pre-tratamiento y eliminación de interferencias. La determinación de capsaicinoides se realiza mediante medidas de absorbancia a 430 nm.

También, relativo al análisis de capsaicinoides, y obteniendo resultados comparables a los obtenidos mediante HPLC, Tsou et al. [128] aplicaron métodos de espectroscopía de infrarrojo cercano (NIR) para la determinación de capsaicinoides. Wang et al. [129] utilizaron técnicas basadas en la espectroscopía de reflectancia difusa visible e infrarrojo cercano, para determinar el contenido de Vitamina C en muestras de chiles. Ignat et al. [130] midieron el contenido en ácido ascórbico en diferentes cultivos de pimientos usando espectroscopia de reflectancia difusa de longitud de onda corta (SWIR) y VIS-NIR.

Baranski et al. [131] determinaron la presencia de luteína, β -caroteno y capsantina, así como los cambios en sus contenidos durante el cultivo de pimientos, mediante espectroscopia NIR con transformada de Fourier-Raman (NIR-FT). El estudio se realizó *in situ*, con medidas de Raman en puntos concretos del fruto y mapeando la superficie de la muestra. Schulz et al. [132] también determinaron el contenido en carotenoides en muestras vegetales, mediante NIR-FT-Raman, consiguiendo la identificación de estos componentes sin tratamiento previo de la muestra.

I.6. Aplicación de métodos de reconocimiento de patrones a la caracterización de alimentos

La autenticación de los alimentos, así como la detección de posibles adulteraciones y fraudes [133] es de vital importancia para los consumidores que buscan productos que garanticen la máxima calidad, autenticidad y pureza, especialmente aquellos que se encuentran dentro del marco de denominación de origen o indicación geográfica protegida. La química analítica cuenta con numerosas herramientas que permiten analizar y determinar los índices o parámetros necesarios para evaluar la calidad, origen o variedad de los productos estudiados. El análisis multivariante (AM), en combinación con las técnicas de reconocimiento de patrones, (PR) constituye un procedimiento muy eficaz en el estudio y caracterización de alimentos [134].

Desde los trabajos de Kowalski sobre la interpretación de datos químicos mediante técnicas de reconocimiento de patrones [135] y sus aplicaciones a la diferenciación de vinos [135, 136], numerosos autores, también, han empleado técnicas quimiométricas para caracterizar diversos tipos de alimentos, y así discriminar orígenes y variedades a partir de datos multivariantes de su composición química.

Luo et al. [137] realizaron un estudio en distintas variedades de manzanas, distintos orígenes y grados de madurez, con el fin de clasificar los distintos tipos, mediante espectroscopía NIR y la aplicación de tres herramientas de PR como el método K vecinos más próximos (KNN), mínimos cuadrados parciales y análisis discriminante (PLS-DA) y el método de ventana móvil junto con PLS-DA (MW-PLS-DA). Los resultados obtenidos mostraron que la mejor técnica para elaborar un modelo que diferencie entre variedades, origen y grado de madurez fue MW-PLS-DA, ya que no solo se consiguen valores de eficacia próximos al 99%, sino que permite la eliminación de interferencias e información ineficiente para la separación. Mangas et al. [138] realizaron un estudio del perfil fenólico de distintas variedades de manzanas para clasificarlas como amargas o no amargas, y para ello analizaron los compuestos fenólicos mediante HPLC en fase inversa y emplearon técnicas

quimiométricas como análisis en componentes principales (PCA), análisis discriminante lineal (LDA) o PLS. Ariana et al. [139] emplearon escaneo multiespectral en los modos reflectancia y fluorescencia para adquirir imágenes de tres variedades de manzanas y detectar defectos en su estructura y apariencia. A partir de las imágenes obtenidas por combinación de filtros desde la región visible hasta el NIR, se crean modelos mediante redes neuronales artificiales (ANN) para clasificar las distintas especies y los posibles defectos que puedan presentar.

Zheng et al. [140] realizaron un estudio del perfil orgánico y metálico de vinagres con denominación de origen protegida para posteriormente construir un modelo aplicando PCA y LDA, que permitiera discriminar los productos con dicha denominación de los que no la poseen. Benito et al. [141] caracterizaron muestras de vinagre de dos DOP como son La Rioja y Jerez, mediante el análisis de 20 variables entre las que se encuentran pH, metales, extracto seco o acidez. Con los datos obtenidos a partir del análisis multivariante, aplicaron herramientas quimiométricas como PCA, ANN, y modelado suave independiente de analogía de clases (SIMCA) para crear modelos que permitieran la clasificación de las muestras con resultados de sensibilidad y selectividad superiores a un 90%. Natera et al. [142], caracterizaron 83 muestras de vinagre según un perfil orgánico, y al conjunto de datos aplicaron análisis clúster (CA) y LDA para diferenciar las muestras según la materia prima (manzana, vino, miel...) y el procesado del vinagre (añada, no añadada), presentando el LDA mejores resultados en ambas clasificaciones. García-Parrilla et al. [143] aplicaron técnicas de reconocimiento de patrones a datos de contenido polifenólico, con objeto de distinguir vinagres de vino procedentes de diferentes vinos y diferentes procesos de acetificación.

El aceite de oliva es un producto cuya materia prima varía mucho en función del origen geográfico de cultivo y producción y, además, presenta una importante relevancia económica en varios países; por ello ha sido, y es, objeto de estudio para numerosos autores. Giansante et al. [144] utilizaron el contenido en ácidos grasos, alcoholes, triterpenos policíclicos y escualeno para la diferenciación de aceites de oliva de origen italiano de diferentes cultivos, combinando técnicas no supervisadas, como PCA, para elegir las variables más discriminantes, y técnicas supervisadas

como LDA, para finalmente crear el modelo de clasificación. Alonso-Salces et al. [145] caracterizaron geográficamente aceites de oliva virgen procedentes de España, Italia, Grecia, Túnez, Turquía y Siria, basándose en los datos extraídos de espectros de ^1H -RMN de la materia insaponificable y tratados mediante LDA, SIMCA, PLS-DA y árboles de clasificación y regresión, (CART). Para la diferenciación de aceite de oliva virgen, aceite de oliva y aceite de orujo, Peña et al. [146] propusieron la aplicación de técnicas de reconocimiento de patrones como CA, PCA, KNN y SIMCA para crear los modelos que permitieran la diferenciación, con valores de eficacia de 97%. Marini et al. [147] analizaron el contenido en ácidos grasos por GC y la composición de esteroides y triglicéridos mediante HPLC, de 40 muestras de aceite de salvado de arroz, con el fin de usar esos datos como variables para la elaboración de modelos mediante LDA y ANN que permitieran clasificar las muestras según su origen. Aplicando PLS y PCA a los datos obtenidos tras la determinación de ácidos grasos de distintos aceites de semillas, es posible agruparlos en función de su año de cosecha [148].

Muchos autores han caracterizado la miel según sus parámetros de control de calidad, como pH, acidez libre y total, contenido metálico, humedad, cenizas, presencia de hidrofurfural, entre otros, con el fin de clasificarlos según origen geográfico. Baroni et al. [149] caracterizaron muestras de miel según los parámetros de calidad y clasificaron muestras de miel procedente del norte y sur de Argentina, mediante KNN, tras reducir las variables y elegir las más discriminantes para el modelo. Yücel y Sultanoglu [150] aplicaron PCA a datos extraídos de análisis de los parámetros de calidad para caracterizar mieles de Turquía. Los mismos autores identificaron también mieles de Hatay usando datos del contenido metálico determinado mediante ICP-OES en combinación con técnicas quimiométricas [151]. Chudzinska y Baralkiewicz [152] estudiaron el perfil mineral, determinando el contenido en Al, B, Ba, Ca, Cd, Cu, K, Mg, Mn, Na, Ni, Pb, Zn de tres tipos de mieles de Polonia, por ICP-MS, y crearon modelos de clasificación usando CA y PCA. Boffo et al. [153] combinaron datos de ^1H -RMN y técnicas quimiométricas como PCA y análisis clúster jerárquico (HCA) para diferenciar entre tipos de mieles de Brasil, según el néctar empleado para la producción de la misma.

Huang et al. [154] desarrollaron un método de análisis del flavor de distintas muestras de café según su perfil volátil, mediante GC-MS. Diferenciaron muestras de café según su flavor, de otras muestras dopadas con aroma de coco, empleando PCA para la diferenciación. También a partir de la fracción volátil, Ribeiro et al. [155] proponen crear modelos mediante PLS, utilizando el algoritmo de selección de predicción ordenado, (OPS), para seleccionar los componentes que más contribuyen a la clasificación, que permitan predecir y clasificar cafés Arábica de Brasil en términos de acidez, amargura, cuerpo del café y claridad, entre otros. Leis Rodríguez de Oliveira et al. [156] compararon recientemente diferentes propiedades sensoriales del café, dependiendo de las distintas formas de cosecha y de tratamientos post cosecha, mediante PCA. Muñiz-Valencia et al. [157] han analizado y determinado el contenido en Ca, Cu, Fe, K, Mg, Mn, Na, Zn en muestras de café Mexicano. En primer lugar se aplicó el test de Kruskal-Wallis para detectar posibles diferencias en la composición mineral, y a continuación PCA para previsualizar la tendencia natural de los datos a diferenciarse. Se aplicó LDA con inclusión sucesiva de variables para corroborar dicha separación, y finalmente se recurrió a modelos creados mediante técnicas de reconocimiento de patrones no lineales como ANN. Un procedimiento similar fue aplicado por Briandet et al. [158] para la determinación de aditivos inusuales en muestras de café instantáneo en términos del perfil orgánico y en concreto de los carbohidratos mediante espectroscopia FTIR. Valentin y Walting [159] determinaron la concentración de 59 elementos mediante análisis por ICP-MS e ICP-AES para determinar el origen de muestras de café de quince países repartidos por los cinco continentes. El estudio estadístico realizado mediante la aplicación de LDA y PCA a los datos de concentraciones permitió una clasificación de las muestras de un 96% de exactitud. Concluyeron que el proceso de tostado, el grado de madurez y la fecha de recolecta no son relevantes en la clasificación.

El té, debido a su alto consumo en todo el mundo [160], ha sido muy estudiado. Valera et al. [161], discriminaron entre dos tipos de té: verde y negro, mediante análisis multivariante y técnicas de reconocimiento de patrones. Consiguieron diferenciar entre tipos a partir de un modelo LDA calculado con variables medidas como polifenoles, aminoácidos totales y xantinas. Alcázar et al. [162] utilizaron el perfil de aminoácidos para diferenciar entre los tipos de té verde,

negro, Oolong, blanco y Pu-erh mediante técnicas quimiométricas como PCA, LDA, KNN y ANN. McKenzie et al. [163] emplearon el contenido metálico de los diferentes tipos de té como descriptores quimiométricos para la diferenciación. En este caso, se emplearon modelos LDA y redes neuronales probabilísticas (PNN). Recientemente, Palacios-Morillo et al. [164] han conseguido diferenciar entre los distintos tipos de té mediante datos de espectros UV-Vis y técnicas de reconocimiento de patrones. Al ser una matriz de datos correlacionados entre sí, se realizó un análisis PCA para la reducción de variables y la obtención de unas nuevas que no estuvieran tan correlacionadas y se trabajó con las coordenadas de las muestras en el espacio de los componentes principales.

En cuanto a las bebidas alcohólicas, existen numerosos autores que han autenticado muestras en términos de denominación de origen y de detección de fraudes. En el caso del vino, el perfil metálico es una excelente fuente de información, ya que las uvas toman metales del suelo y, por tanto, la correlación entre el contenido y la zona de origen es muy razonable [165]. El contenido en metales ha sido estudiado por Latorre et al. [166], quienes determinaron la concentración de Li, Na, Rb, entre otros, mediante espectroscopia de emisión atómica de llama (FAES), con el objetivo de discriminar las distintas denominaciones de orígenes y adulteraciones de vinos gallegos, haciendo uso de PCA, LDA, CA, KNN y SIMCA. Además de FAES, Frías et al. [167] usaron FAAS para la determinación del contenido metálico de vinos canarios y emplearon los datos para la clasificación de éstos mediante PCA, CA, LDA y SIMCA. Mediante ICP-OES, González et al. [168] determinaron el contenido mineral, y aplicando CA y PCA diferenciaron muestras de los orígenes valencianos de Utiel-Requena y Jumilla. Tras aplicar PCA y DA a los datos de concentraciones de elementos como Na, K, Al, As, Cu, Rb, Sr, entre otros, obtenidos mediante ICP-OES e ICP-MS, Sporkova y Suchánek [169] identificaron orígenes de vinos de la República Checa. Jurado et al. [170] clasificaron vinos blancos de cuatro DO españolas de acuerdo con el perfil elemental de las muestras, utilizando máquinas de vectores soporte (SVM). La fracción orgánica en las muestras de vino también se ha utilizado muy frecuentemente. Por un lado, el color se ha empleado como parámetro diferenciador por varios autores. Urbano et al. [171] usaron espectroscopia UV-Vis para la

diferenciación de vinos de La Mancha, según el tipo de uva, el origen y la añada. Para ello aplicaron PCA como técnica de previsualización, y a continuación, un modelo de clasificación mediante SIMCA. Cazzolino et al. [172] emplearon datos de espectroscopía Vis y NIR en combinación con PCA, regresión en componentes principales (PCR) y DA-PLS para la clasificación de vinos australianos. Fernández y Agosin [173] cuantificaron la cantidad de taninos presentes en muestras de vinos tintos ya que estos compuestos son indicadores de la calidad de los vinos. El análisis químico se realizó mediante FT-NIR, y los datos se trataron con PLS para crear un modelo capaz de predecir la cantidad de taninos. Jurado et al. [174] y Villiers et al. [175] aplicaron GC y HPLC, respectivamente, junto con técnicas quimiométricas para diferenciar tipos de vinos. Alcázar et al. [176] diferenciaron entre tipos de cerveza usando como descriptores el contenido de cloruros, fosfatos, sulfatos aminoácidos totales y pH. Emplearon PCA para previsualizar la tendencia natural de los datos a separarse, y LDA para crear el modelo que diferencia entre cerveza Lager, Pilsner y de bajo contenido alcohólico. Los mismos autores [177] diferenciaron el origen geográfico de cervezas mediante el cálculo de modelos LDA a partir de parámetros químicos como pH, polifenoles totales y el contenido metálico de las muestras. Biancolillo et al. [178] emplearon varias técnicas analíticas distintas: termogravimetría, espectroscopía de NIR, MIR y UV-Vis para caracterizar y garantizar la autenticidad de una cerveza de alta calidad, empleando como herramientas quimiométricas SIMCA y DA-PLS. Existen estudios de diferenciación para otras bebidas como tequila [179-188], anís o raki [181,182], ron, brandy, whisky y vodka [183].

Este tipo de estudios es muy útil para la detección de fraudes. Moore et al. [184] confeccionaron una base de datos recogiendo los fraudes más frecuentes en la composición de alimentos. Los más habituales son los detectados en leche, aceite de oliva, miel, café y azafrán.

De la Fuente y Juárez. [185] consideran la adulteración de la leche particularmente peligrosa, ya que es materia base para la industria láctea, la cual genera numerosos derivados. Santos et al. [186] proponen la detección de fraudes en leche mediante medidas por microespectroscopia de Infrarrojo medio (MIR) de los

principales adulterantes como urea, suero de leche, peróxido de hidrógeno, orina sintética, orina y leche sintética. Mediante la aplicación de SIMCA a los datos obtenidos consiguieron discriminar entre la muestra de control y la adulterada con suero, urea, urea sintética, leche sintética y peróxido de hidrógeno, mientras que aplicando regresión por mínimos cuadrados parciales (PLSR) se obtienen funciones que distinguen la muestra de control de aquellas dopadas con suero, orina sintética, urea leche sintética y peróxido de hidrógeno. Otros autores como Bassbasi et al. [187] y Botros et al. [188] han determinado fraudes y adulteraciones en la leche empleando espectroscopía IR y técnicas quimiométricas.

Mediante datos extraídos de análisis por IR y NIR, Downey et al [189] y Kelly et al. [190] discriminaron entre muestras de miel sin adulterar y muestras adulteradas mediante la incorporación de azúcares añadidos. Para ello emplearon técnicas de reconocimiento de patrones como KNN, PLS-DA y SIMCA. Cordella et al. [191] aplicaron LDA a datos procedentes del análisis de 13 azúcares mediante HPLC para la discriminación de muestras adulteradas mediante la adición de azúcares y la modificación de la alimentación de las abejas. López et al. [192], determinaron mediante técnicas de reconocimiento de patrones parámetros relacionados con el control de calidad de mieles como, por ejemplo, humedad, cenizas, materia insoluble, azúcares reductores, acidez total y libre o hidroximetilfurfural, entre otros. Para la detección de edulcorante como elemento adulterante en mieles, Zhu et al. [193] analizaron la composición de la miel mediante NIR, reduciendo la matriz de datos mediante PCA y posteriormente crearon cinco modelos de clasificación para discriminar entre muestras adulteradas y sin adulterar mediante máquinas de vectores soporte por mínimos cuadrados (LS-SVM), SVM, ANN, LDA y KNN. Ordoudi et al. [194] combinaron datos extraídos de FT-IR, complementados con análisis por HPLC-DAD, con PCA para detectar adulteraciones en muestras de azafrán.

Debido a la importancia que tienen las especias, hierbas aromáticas y condimentos en todo el mundo, el estudio de diferenciación de éstas también se ha llevado a cabo por varios autores. Palacios-Morillo et al. [195] han diferenciado denominaciones de origen de pimentón elaborado en España a partir del contenido

metálico. Korel et al. [196], diferenciaron distintos tipos de pimentones turcos mediante DA utilizando la concentración total de capsaicinoides. Yudthavorasit et al. [197] diferenciaron muestras de jengibre según el origen de procedencia, mediante el análisis por HPLC de gingerol y otros compuestos relacionados y técnicas quimiométricas como PCA, HCA y LDA, concluyendo que era posible asegurar la procedencia de las muestras de jengibre mediante este procedimiento. Karadas y Kara [198] evaluaron, mediante análisis quimiométrico, la relación existente entre hierbas y especias en términos de concentración de elementos metálicos. Así, mediante PCA y CA se agruparon las muestras de pimienta negra, romero, pimentón dulce y picante, cúrcuma, canela, comino, menta y tomillo en cinco grupos dependiendo del contenido mayoritario de elementos como Ba, Co, Cr, Fe, Mn, Ni, Sr y se comprobó mediante LDA que la clasificación en estos grupos era correcta. D'Archivio et al. [199] utilizaron el contenido mineral y de elementos traza determinado mediante ICP-MS para clasificar azafranes de distintos orígenes de Italia. Para ello, identificaron los elementos más discriminantes y generaron modelos de clasificación mediante LDA. Anastakasi et al. [200] llevaron a cabo la geográfica de muestras de azafrán procedente de Irán, Italia, España y Grecia, a partir de los datos analíticos de la fracción volátil, analizada mediante GC-MS FID, y en combinación con análisis discriminante canónico (CDA). Así mismo, estos autores [201], utilizaron espectroscopía de IR medio para diferenciar muestras de azafrán de cuatro orígenes diferentes, empleando para ello los datos procedentes de AM y PCA.

I.7. El color

I.7.1 Breve introducción histórica

El color, indudablemente, ha estado presente en la vida del hombre desde sus inicios, y prueba de ello son las pinturas rupestres o las cerámicas egipcias [202]. Aristóteles ya mostró interés hacia el color escribiendo su obra denominada *De coloribus*, donde manifestaba que el color era una propiedad de la luz. Newton, en su obra *On colour* (sobre el color) aporta los primeros conocimientos afirmando que un observador puede duplicar el efecto de un estímulo de color combinando la luz de tres fuentes primarias de color, denominadas valores triestímulo. Goethe rebatió la

teoría de Newton analizando el color como algo subjetivo y propio del observador [203]. El matemático alemán Hermann Grassmann [204], apoyándose en la aseveración de Newton, estudió los experimentos previos y enunció unas leyes sobre la mezcla aditiva del color, mostrando que, efectivamente, cualquier color puede expresarse como suma de tres colores primarios. Maxwell, apoyándose en las ecuaciones de Gassmann, fue el primero en determinar cuál es la cantidad de componentes primarios tricromáticos que igualan las diversas radiaciones puras del espectro [205]. En 1924 se crea la *Commission Internationale de l'Eclairage* (CIE), que establece las condiciones experimentales de referencia y estandariza observadores e iluminantes para inferirle al concepto y a los cálculos de color un uso de carácter internacional.

A lo largo de la historia, el termino color se ha usado de manera diferente dependiendo del contexto o del interés del momento y del usuario, generando, en algunos casos, cierta confusión con los términos, expresiones o vocabulario. En la figura I.7 se presenta un esquema donde se resume una forma de entender el estudio del color y su medida

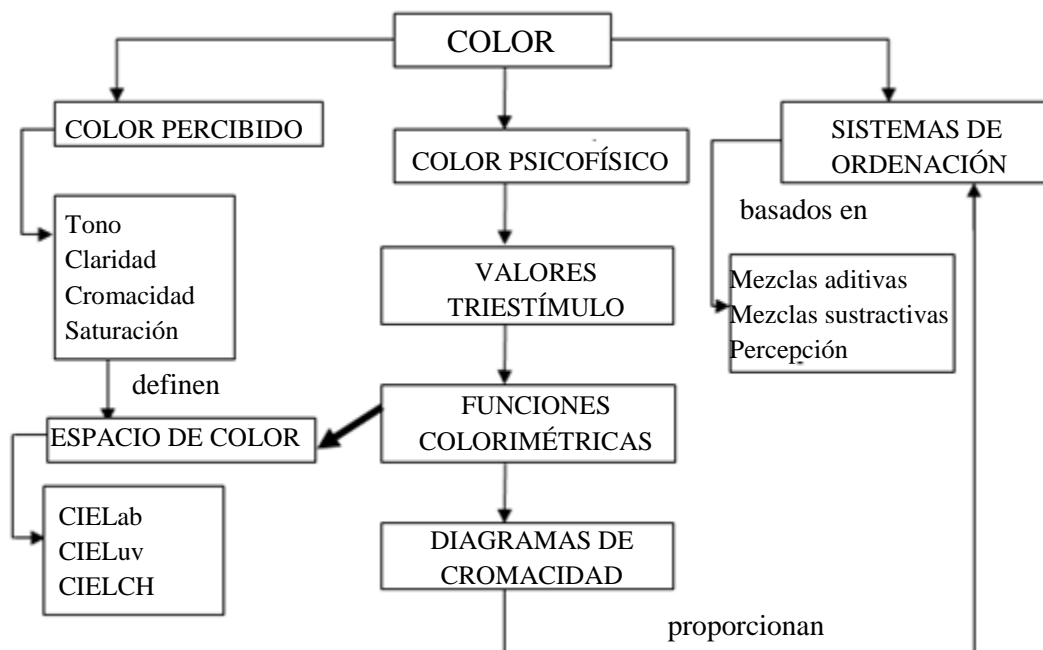


Figura I.7. Conceptos relacionados con el color.

I.7.2 Definición del color

El color es un atributo de la visión y una característica de la luz. En función del interés del usuario, el término color se puede definir de una manera u otra. Por ello, la norma UNE y la CIE definen dos conceptos: color percibido y color psicofísico.

El color percibido es un atributo de la percepción visual mediante el cual un observador puede diferenciar entre dos objetos del mismo tamaño, forma y textura basándose en las diferencias en la composición espectral (combinación cualquiera de elementos cromáticos y acromáticos) de las radiaciones relacionadas con la observación.

El color psicofísico es la característica de la radiación visible que permite al observador distinguir diferencias entre dos objetos de igual forma y estructura.

El color percibido es el que caracteriza la visión del color en términos de tono, brillo o luminosidad y saturación. El tono es el atributo de una sensación visual según el cual una superficie parece ser semejante a uno de los colores percibidos, rojo, amarillo, verde o azul, o a una combinación de dos de ellos. Existen colores percibidos sin tono que se denominan colores acromáticos y colores percibidos con tono, referidos como colores cromáticos. El brillo o luminosidad es el atributo de una sensación visual según el cual una superficie parece emitir más o menos luz. En el caso del color podemos hablar de claridad como la luminosidad de una superficie juzgada en relación a la luminosidad de una superficie que parece blanca o posee una reflectancia elevada y está iluminada de idéntico modo. La cromacidad o contenido de color es el atributo de una sensación visual según el cual una superficie parece presentar un color percibido más o menos cromático. Estos tres términos, luminosidad, tono y contenido de color son las tres características básicas del color percibido con superficies emisoras [206].

Existen además otros términos, en especial si nos referimos a colores dependientes y superficies receptoras. Haciendo un paralelismo con los términos empleados para el color percibido con superficies emisoras, se define la claridad

como luminosidad de una superficie evaluada en relación a la luminosidad de otra superficie, igualmente iluminada, o que parece blanca o altamente difusora. Para definir el contenido de color, en este caso se necesitan dos términos: la saturación, que es la cromacidad de una superficie evaluada en proporción a su luminosidad, y el croma, que es el contenido de color de una superficie evaluado en proporción a la luminosidad de otra superficie, igualmente iluminada, que parece blanca o altamente difusora. Otros conceptos importantes a la hora de definir el color percibido son la cromaticidad (percibida), que es el atributo de la sensación visual formada por el tono y la saturación, el color (percibido) luminoso, que se define como el color percibido de una superficie que emite luz como una fuente de luz, o que refleja especularmente la luz, y el color (percibido) no luminoso, como color percibido de una superficie que transmite o refleja difusamente la luz. A la hora de hablar de los conceptos relacionados con el color, es importante aclarar que cuando se trata de colores de sólidos, nos referimos a la luz reflejada (no absorbida), mientras que cuando hablamos del color de líquidos, y disoluciones, hay que referirse a la luz transmitida.

I.7.3 Medidas del color

El color psicofísico está relacionado con la *colorimetría*, que es el conjunto de métodos para evaluar y medir el color, lo cual implica la asignación de números que relacionen el fenómeno psicológico (color) con el fenómeno físico (longitud de onda etc.) que provoca la percepción. Todas las medidas del color reciben el nombre de medidas psicofísicas.

Las medidas del color tienen su base en la evaluación de un color en función de estímulos primarios o normalizados. El principio enunciado por Newton [207] y verificado por Young [208], basa la especificación de estímulos de color en la igualación de colores por mezcla aditiva de cantidades adecuadas de tres estímulos de color de referencia, que constituyen un sistema tricromático. Estas cantidades de los estímulos de referencia que en un sistema tricromático dado son necesarios para igualar el estímulo considerado, se denominan valores triestímulo. Existen diferentes

formas de representar los valores triestímulo que varían en función del sistema tricromático. Sobre la base de esta propiedad se construye la colorimetría que conocemos hoy en día.

H. Grassmann introdujo las dimensiones del espacio de color a través de tres magnitudes independientes, tono, intensidad del color e intensidad del blanco. Los estudios de Grassmann [204] parten de un artículo de Helmholtz [209] y se basan en la hipótesis de Newton [210] según la cual para producir el blanco son necesarios tres colores espectrales. Por ello Grassmann propone sus cuatro postulados:

- I. Una sensación de color se especifica completamente con tres magnitudes: tono, brillo del color y brillo del blanco.
- II. Si una luz varía de forma continua, la sensación de color de la mezcla aditiva con otra luz fija también varía de forma continua.
- III. El resultado de una mezcla aditiva de colores depende sólo de su aspecto y no de su composición física.
- IV. La intensidad luminosa de una mezcla aditiva de colores es la suma de la intensidad luminosa de los colores mezclados.

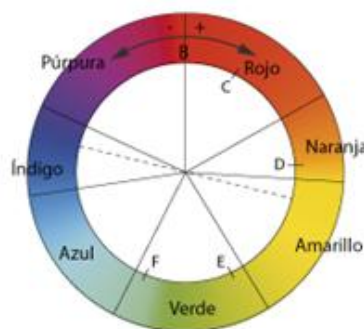


Figura I.8. Rueda de color de Grassmann.

Estos postulados destacaron, sobre todo, por dos aspectos. En primer lugar por el concepto de valores triestímulo, que más adelante se concluirá que esos tres vectores pueden ser también los colores primarios. Por otro lado, Grassmann introduce el tema de la continuidad a la hora de plasmar la rueda de color (figura I.8), a diferencia de Newton que proponía una división discreta del espectro en siete colores.

El sistema triestímulo más empleado hoy en día, es el que propone la CIE.

▪ *CIE*

La CIE propone el uso de un campo de visión global, nivel visual fotópico (luminancia superior a 10^{-1} candelas/ pie^2) y como estímulos de referencia se debían usar radiaciones espectrales de 435,8, 546,1 y 700 nm [211]. W. D. Wright [212] y J. Guild [213], con la ayuda de diez y siete observadores respectivamente, establecieron valores experimentales de las magnitudes de los primarios que igualaban los colores espectrales. Se normalizaron los resultados para hacer comparables las funciones entre sí.

- *Observador CIE RGB*

Se eligen tres estímulos correspondientes con tres colores primarios rojo (R), verde (G) y azul (B), que se pueden representar por λ_R , λ_G y λ_B . En el sistema CIE 1931 se toman como primarios el 435,8 nm (violeta), el 546,1 (verde) y el 700 (rojo). Para igualar los estímulos a una radiación dada se aplica la Ecuación 1.1:

$$a(\lambda) = b(B) + c(G) + d(R) \quad (1.1)$$

y con el fin de normalizar, para los cocientes se establecen los siguientes coeficientes:

$$(\lambda) = \frac{b}{a}(B) + \frac{c}{a}(G) + \frac{d}{a}(R) \quad (1.2)$$

Estableciendo que

$$\frac{b}{a} = \bar{b}; \frac{c}{a} = \bar{g}; \frac{d}{a} = \bar{r} \quad (1.3)$$

La ecuación queda como,

$$(\lambda) = \bar{b}(B) + \bar{g}(G) + \bar{r}(R) \quad (1.4)$$

Uno de los problemas que presenta este sistema de coordenadas es que las funciones colorimétricas $r(\lambda)$, $b(\lambda)$ y $g(\lambda)$ muestran valores negativos, y el manejo de cantidades negativas para referirse al color no es lo más intuitivo para el usuario. Por ello surge el observador CIE XYZ.

- *Observador CIE XYZ*

Este observador queda definido por tres funciones derivadas de la sensibilidad del ojo humano hacia tres estímulos primarios. En este caso no se trata de tres colores reales, sino de estímulos denominados X, Y, Z derivados del rojo (R), verde (G) y azul (B) mediante una serie de ecuaciones. Estas funciones colorimétricas son aplicables a campos de visión de 4°. Posteriormente, se recomendó el observador suplementario CIE 1964 para aquellos casos en que se desee una correlación con la visión de objetos con un ángulo mayor de 4°.

- *Observador suplementario CIE XYZ*

Esta serie de valores triestímulo patrón, se crean porque el sistema colorimétrico patrón CIE 1931 sólo es aplicable a ciertos campos visuales con ángulos de observación de hasta 4°. En la práctica, los campos de visión que se emplean son mayores, por lo que la CIE en 1959 publicó dichos valores espectrales basándose en trabajos y experimentos anteriores. Así se definió en *Observador Colorimétrico UNE 4 (CIE 1964)*, al que también se llamó Observador Suplementario y Observador 10°, ya que las medidas se obtuvieron empleando un ángulo de observación de 10° [214].

En ocasiones la representación tridimensional es complicada, y por ello se prefiere una representación en un plano definido por unas coordenadas de cromaticidad que son cada valor triestímulo dividido por su suma.

I.7.4 Ordenación del color

Los sistemas de ordenación de color se hacen necesarios, ya que la percepción del color es algo subjetivo e intercambiar información sobre el color es

complicado. Se ordenan los colores de manera metodológica, racional, y a partir de un sistema de referencia normalizado que facilite la percepción rápida de la relación que existe entre colores y la posición de uno con respecto al resto. Los sistemas de ordenación se basan en las mezclas de colores y por ello se pueden clasificar en aditivos, partitivos y sustractivos y sistemas basados en la percepción.

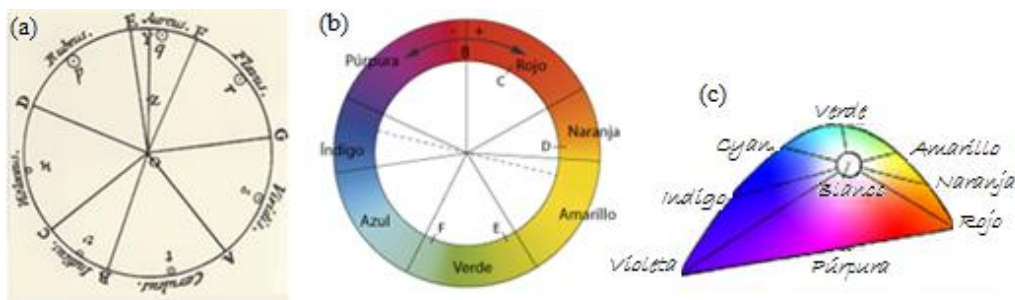


Figura I.9. Esquema original de color. (a) Newton; (b) Grassmann; (c) Helmholtz.

Dentro de los sistemas de ordenación aditivos, Newton fue el primero en ordenar los colores realizando diagramas del espectro visible en el que las luces de color están en forma circular. El rojo, verde y azul eran tomados como colores primarios, y los complementarios amarillo, naranja, índigo y violeta alrededor del círculo. Estos siete colores, originariamente sin distribución uniforme, como puede verse en la figura I.9., componen el círculo cromático de Newton. Grassmann duplicó el círculo de Newton, pero tampoco incluyó el púrpura, aunque enunció que era el complementario del verde. Helmholtz propuso que los colores del espectro no se disponían en torno a un círculo, sino sobre un diagrama de forma particular en el que actúa la ley del baricentro y que constituyó una primera revisión importante del diagrama de los colores espectrales de Newton y Grassmann.

Los sistemas gráficos, relacionados con los medios informáticos, ofrecen modelos de ordenación del color basados en sistemas aditivos, como puede ser el modelo RGB. Existe cierta controversia a la hora de hablar de modelos RGB, ya que algunos autores consideran el RGB como un sistema de ordenación [211] y otros como un espacio de color definido por tres vectores [215]. Entendido como sistema de ordenación, se basa en la teoría tricromática donde los colores se definen mediante

mezcla de los colores primarios, rojo, verde, azul (Red, Green, Blue, RGB) por adición (Figura I.10a).

Cuando se habla de los sistemas de ordenación sustractivos, se hace referencia sobre todo a los atlas y catálogos de ordenación del color como el manual publicado por Hicketier [216] o el Atlas de Color de Villalobos y Villalobos [217]. Dentro de los sistemas sustractivos cabe destacar también el sistema propuesto por Ostwald [218], basado en colores marginales y sus combinaciones. Otros sistemas de ordenación sustractivos de tipo gráfico, es decir, basados en métodos informáticos, son los modelos CMY y CMYK (Figura I.10b).

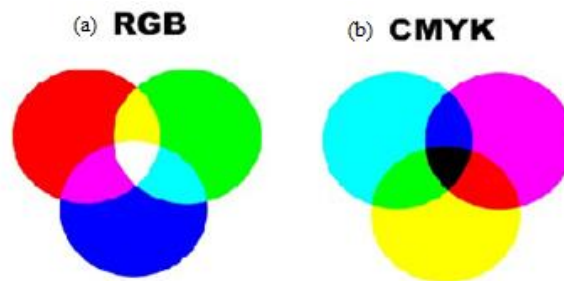


Figura I.10. (a) modelo RGB, (b) modelo CMYK.

Ostwald crea un espacio de color basado en la longitud de onda dominante, pureza y luminancia, representándolos con los valores de matiz, saturación y brillo. Los colores se pueden ver en el corte del plano horizontal del sólido de colores. El círculo se divide en ocho sectores diferenciados: amarillo, naranja, rojo, púrpura, azul ultramar, turquesa, verde mar y verde hoja. Cada uno de estos sectores está dividido en tres partes, dando un total de 24 tonos para todo el círculo. El número variable de cada división se indica por la letra N que va de 1 a 24. Dividiendo el doble sólido por un plano que contenga al eje común, se obtienen dos triángulos cuyos tonos son complementarios, cada uno de los cuales está dividido en 28 colores de un mismo tono. A estos triángulos Ostwald los llama triángulos monocromáticos. El eje vertical del doble cono está dividido en 8 pasos perceptualmente iguales que van desde el blanco al negro. La saturación crece desde el eje central hacia afuera. Los tonos están distribuidos a lo largo del círculo central de tal manera que cada color tiene diametralmente opuesto a su complementario. De esta manera, el punto en el espacio de color Ostwald, está representado por la ecuación 1.5:

$$z + k + w = 1 \quad (1.5)$$

Donde z indica la cantidad de blanco, k , la cantidad de color y w , la cantidad de negro.

Las ventajas de este sistema de ordenación de colores es la simplicidad, pero su teoría no era correcta, ya que evaluación de la reflectancia espectral en base a la existencia de colores puros, como él define, no es cierta.

Los sistemas de ordenación basados en la percepción son, principalmente, escala de grises Munsell, DIN, NCS, UNE y el modelo HSV. El sistema de escala de grises se basa en construir una escala visual en cinco etapas a partir de blancos y negros. El sistema Munsell [219], creado por Albert Henry Munsell, consiste en una órbita donde el ecuador lo forman una banda de colores y los ejes son una escala de valores donde el blanco es el norte y el negro el sur. Los ejes se denominan tono, claridad y croma (figura I.11). El sistema DIN, es una norma alemana basada en un sólido de color espaciado colorimétricamente donde las tres dimensiones son similares al tono, croma y claridad de Munsell. El sistema NCS-UNE, *Natural Color System*, está basado en las percepciones de los seis colores: amarillo, rojo, azul, verde, blanco y negro; todos los demás colores son mezcla de los anteriores. El modelo HSV (figura I.12a) es parte de un sistema gráfico, por lo tanto de índole informática, donde las variables que definen al color son tono (H), saturación (S) y luminosidad (L). En este caso la geometría del modelo es un cono cuya base superior es el círculo cromático.

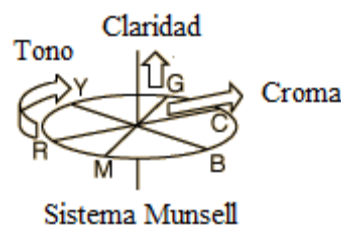


Figura I.11. Sistema Munsell.

Los sistemas descriptivos dentro de los basados en la percepción son los nombres de color, diccionarios de color, el *colour index* y el modelo HSL (figura

I.12b). Este último es un sistema gráfico que también responde a una geometría cónica, en este caso doble cónica, y presenta ciertas analogías con el sistema Munsell.

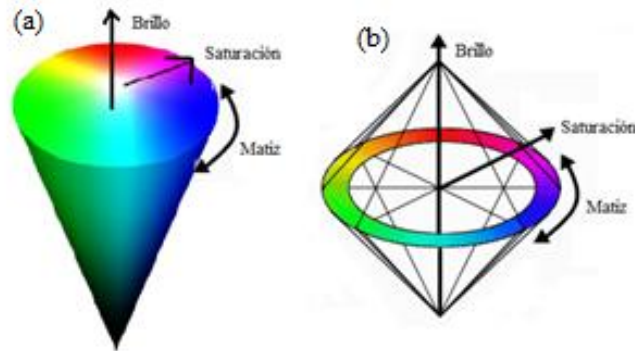


Figura I.12. (a) modelo cónico HSV; (b) modelo doble cónico HSL.

La forma definitiva del diagrama, que acabaría llamándose diagrama de cromaticidad, la presentaría la CIE en 1931, y aunque no es estrictamente un método de ordenación del color, es el medio más conveniente para representar las relaciones entre colores.

I.7.5 Espacios de color

Una vez se establece el observador patrón CIE 1931, con los valores triestímulo X, Y, Z, es posible introducir muchas coordenadas calculadas a partir de expresiones matemáticas. Estas coordenadas sitúan cada color en un espacio determinado, que tendrá tantas dimensiones como coordenadas se necesiten. Por ello, cuando se eligen las coordenadas colorimétricas, queda definido un espacio de color. Un espacio de color es la representación geométrica de los colores en el espacio, generalmente de tres dimensiones, mediante vectores que tienen por origen el punto representativo del estímulo acromático. Una de las propiedades más importantes que se le va exigir a cualquier espacio de color utilizado para nombrar colores es la uniformidad, es decir, que todas las diferencias de color igualmente percibidas se definan en el espacio por las mismas distancias.

Como se ha mencionado, un espacio puede ser descrito por diferentes sistemas de coordenadas colorimétricas y valores triestímulo, como por ejemplo, el sistema Munsell, Ostwald, RGB o CIE. El sistema Munsell usa el tono o matiz, el brillo y el croma, mientras que el Ostwald describe el espacio con las coordenadas pureza, luminancia y longitud de onda dominante. Usar los colores primarios RGB como vectores, también es útil para definir un espacio de color. En la figura I.13. se representan estos espacios. El sistema de coordenadas más común y recomendado es el descrito por la CIE, y por ello se prestará especial atención a éste. En un primer momento se empleó el cálculo de las coordenadas XYZ, pero en 1976 la CIE perfeccionó su sistema de medida e instauró y recomendó el uso de las coordenadas L^* , a^* , b^* , creando así el espacio CIELab.

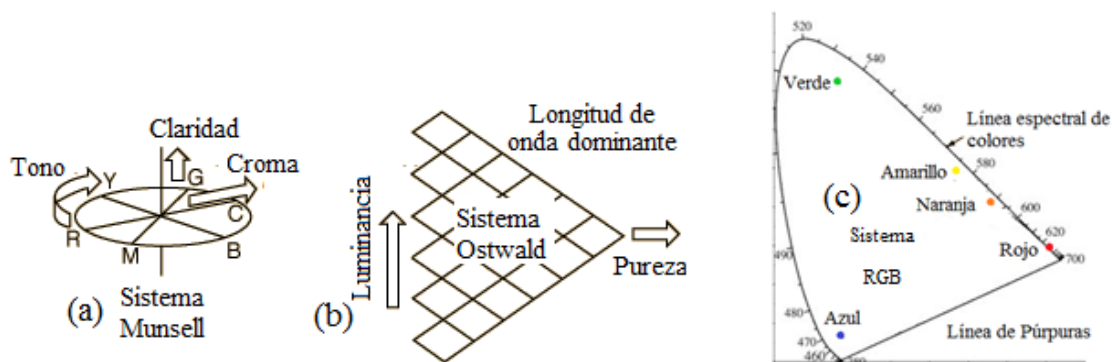


Figura I.13. Espacios de color basados en sistemas (a) Munsell, (b) Ostwald, (c) RGB

El espacio CIELab es el recomendado en numerosos casos por la CIE y otros autores [220], aunque en algunas aplicaciones se recomienda el uso de otros espacios de color como el CIELUV o el espacio CIEUVW. En general, la disyuntiva que pueda presentar el uso de un espacio de color u otro viene dado por el problema que supone definir un espacio verdaderamente uniforme. Existen otros espacios de color en la bibliografía como son el LABHNU [221] o el espacio *zeta* de Sanderson-Milner [222].

I.7.5.1 Espacio CIELab

El espacio CIELab es usado para describir todos los colores que percibe el ojo humano. Los vectores que definen el espacio tridimensional son la claridad, L^*

($L^*=0$ negro y $L^*=100$ blanco), la posición entre el magenta y verde ($a^*<0$ verde, $a^*>0$ magenta) y la posición entre el amarillo y azul ($b^*<0$ azul, $b^*>0$ amarillo).

A continuación explicaremos la relación matemática entre XYZ y las coordenadas $L^*a^*b^*$, partiendo de las coordenadas X, Y, Z que se calculan según las siguientes expresiones:

$$X = K \cdot \sum_{\lambda} T_{(\lambda)} S_{(\lambda)} \bar{X}_{10(\lambda)} \Delta_{(\lambda)} \quad (1.6)$$

$$Y = K \cdot \sum_{\lambda} T_{(\lambda)} S_{(\lambda)} \bar{Y}_{10(\lambda)} \Delta_{(\lambda)} \quad (1.7)$$

$$Z = K \cdot \sum_{\lambda} T_{(\lambda)} S_{(\lambda)} \bar{Z}_{10(\lambda)} \Delta_{(\lambda)} \quad (1.8)$$

donde $T_{(\lambda)}$ es la medida de transmitancia de la muestra para cada longitud de onda, $\Delta_{(\lambda)}$ el paso de longitud de onda elegido para registrar el espectro de absorbancia en el rango de medida, en este caso 1 nm, $S_{(\lambda)}$ es un coeficiente que depende del iluminante usado, $\bar{X}_{10(\lambda)}$, $\bar{Y}_{10(\lambda)}$ y $\bar{Z}_{10(\lambda)}$ son coeficientes que dependen del observador considerado y que están recogidos en las tablas publicadas por la CIE [223], y K es una constante de normalización que se calcula según [224]:

$$K = \frac{100}{\sum_{\lambda} S_{(\lambda)} \bar{Y}_{10(\lambda)} \Delta_{(\lambda)}} \quad (1.9)$$

Una vez obtenidos los valores triestímulo, es posible calcular a partir de ellos las coordenadas L^* , a^* y b^* , que definen el espacio CIELab. La claridad, L^* se calcula según la ecuación 1.10 y en los casos en que $Y/\bar{Y}_n \leq 0,008856$ hay que incluir la modificación de Pauli para evitar la pérdida de uniformidad. la ecuación utilizada en este caso para el cálculo de L^* se muestra en la ecuación 1.11. Los valores \bar{X}_n , \bar{Y}_n y \bar{Z}_n , como ya se ha explicado, se corresponden con un estímulo blanco de referencia que dependerá del iluminante elegido.

$$L^* = 116 \cdot \left(\frac{Y}{\bar{Y}_n} \right)^{1/3} - 16 \quad (1.10)$$

$$L^* = 903,3 \cdot \frac{Y}{\bar{Y}_n} \quad (1.11)$$

Las coordenadas a^* y b^* forman un plano perpendicular con la coordenada L^* (figura I.14.). Para el cálculo de a^* y b^* se emplean las siguientes ecuaciones:

$$a^* = 500 \cdot f(1) - f(2) \quad (1.12)$$

$$b^* = 200 \cdot f(2) - f(3) \quad (1.13)$$

donde,

$$f(1) = \left(f \left(\frac{X}{\bar{X}_n} \right) \right) \quad (1.12)$$

$$f(2) = \left(f \left(\frac{Y}{\bar{Y}_n} \right) \right) \quad (1.15)$$

$$f(3) = \left(f \left(\frac{Z}{\bar{Z}_n} \right) \right) \quad (1.16)$$

Las funciones $f(i)$ son función de los valores triestímulo, que dependen de la claridad de la muestra y se relacionan con las siguientes variables:

$$G(1) = \frac{X}{\bar{X}_n} \quad (1.17)$$

$$G(2) = \frac{Y}{\bar{Y}_n} \quad (1.18)$$

$$G(3) = \frac{Z}{\bar{Z}_n} \quad (1.19)$$

Cuando

$$G(i) > 0,000856 \Rightarrow f(i) = G(i)^{1/3} \quad (1.20)$$

y si

$$G(i) < 0,000856 \Rightarrow f(i) = 7,787 \cdot G(i)^{1/3} + \frac{16}{116} \quad (1.21)$$

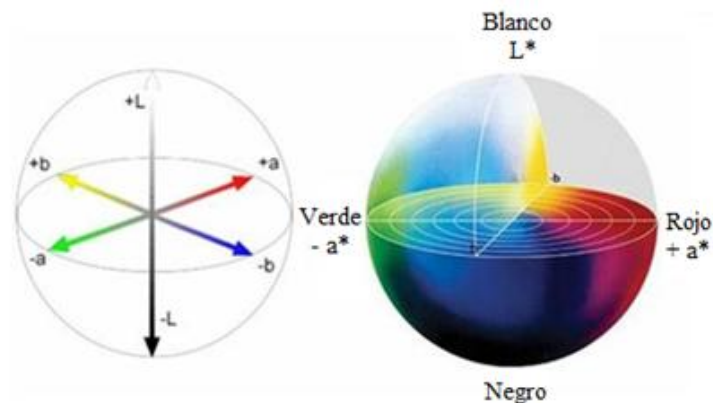


Figura I.14. Espacio CIELab.

I.7.6 El color en pimientos y pimentones

Los responsables del color en los pimientos y el pimentón son los carotenoides, que son pigmentos orgánicos del grupo de los isoprenoides, que se encuentran de forma natural en plantas y otros organismos fotosintéticos como algas, algunas clases de hongos y bacterias. Se conoce la existencia de más de 700 compuestos pertenecientes a este grupo. Los carotenoides son el grupo más representativo de los tetraterpenos, compuestos que se caracterizan normalmente por tener una estructura con 40 átomos de carbono. Estos átomos de carbono se encuentran ordenados formando cadenas poliénicas conjugadas, en ocasiones terminadas en anillos de carbono. A los carotenoides que contienen átomos de oxígeno se les conoce más específicamente como xantofilas. Los restantes constituyen el grupo de los llamados carotenos. La concentración y la composición de carotenoides son responsables de los diversos colores observados en los pimientos [225]. Así, la clorofila contribuye al color verde [226], α - y β -caroteno, zeaxantina, luteína and β -criptoxantina lo hacen al color amarillo-naranja [225], y el pigmento rojo es el resultado de la presencia de capsantina y capsorubina [227]. Esta variedad en los colores se encuentra directamente relacionada con su estructura. Mientras el número de enlaces dobles conjugados aumenta, la longitud de onda de la luz absorbida también lo hace, dando al compuesto una apariencia más rojiza.

En el caso de los frutos del género *Capsicum* y del pimentón, el color toma una especial relevancia, ya que se emplea como colorante y, además, como agente antioxidante [228]. Es evidente que la apariencia física y la presentación de alimentos juegan un papel decisivo en la aceptación del consumidor [229], siendo el color una de las características organolépticas más cruciales, ya que afecta de forma inmediata a la apariencia del producto [230]. Si se hace referencia al extracto o el pimentón, color y calidad del producto están relacionados en el sentido de que la calidad aumenta al aumentar la capacidad colorante [231].

Los pimientos frescos son de los vegetales con una concentración mayor de vitamina C [232] y los efectos beneficiosos para la salud que presentan los pimientos se atribuyen, en parte, a su contenido en dichos carotenoides. Hay estudios que revelan que el consumo de estos compuestos liposolubles ofrece protección contra ciertos cánceres, previenen úlceras gástricas, estimulan el sistema inmunológico, y protegen contra enfermedades cardiovasculares o de cataratas [233].

Por ello, la medida del color asegura la calidad del producto, que es, por otra parte, la base para las transacciones comerciales [234, 235]. El contenido de pigmentos de pimientos se ha estudiado en profundidad para determinar los cambios que sufre el color en el proceso de almacenamiento [238] y tras el secado del producto fresco [237].

Para la medida del color a nivel industrial, la simple observación visual no es adecuada, ya que no es fácilmente reproducible, al depender de la apreciación personal del observador, de la luz incidente o de la humedad. Hasta la fecha, existen varios métodos para la medida del color en general y, en particular, para el pimiento y el pimentón. El método más frecuente es el propuesto por la Asociación Americana para el Comercio de Especies (ASTA), que consiste en una medida de absorbancia a una longitud de onda de 460 nm de una muestra extraída en acetona [79]. Es la medida utilizada como criterio de calidad por el Consejo Regulador de DOP de pimentón de Murcia y La Vera [26, 35]. Es una prueba efectiva en la medida del color pero cuenta, principalmente, con el inconveniente de que no es capaz de detectar adulteraciones debidas a la adición de otros colorantes [236]. Existen otros

métodos para la medida de color cómo el cálculo de los parámetros L^* , a^* y b^* del espacio CIELab de color. En la bibliografía se encuentran procedimientos para la medida por ambos métodos y autores que afirman que no son intercambiables sino complementarios.

Lannes et al. [239] estudiaron el color de 49 muestras de *Capsicum chinense* mediante el método ASTA observando que las muestras que presentaban valores ASTA entre 0-99 eran frutos amarillos en su mayoría, y también anaranjados. En el rango de valores comprendido entre 100-199 unidades ASTA encontraron que, aproximadamente la mitad de los frutos eran anaranjados y la otra mitad rojos. Para valores superiores a 200 unidades ASTA, todos los frutos eran rojos.

Krajayklang et al. [240] estudiaron mediante el método CIELab las diferencias de colores en pimientos en distintos estadios de la cosecha, así como la influencia de una adición exógena de etileno a los pimientos ya cosechados. También midieron el color del extracto seco del fruto por el método ASTA. Los resultados obtenidos mostraron que existen diferencias en los colores en función del momento de cosecha, y que para la detección de color del pimiento seco y molido, el método CIELab no es una alternativa al método ASTA ya que no refleja con exactitud los cambios de color producidos. También en esta línea, Mínguez-Mosquera et al. [234], estudiaron el color en varias muestras del género *Capsicum*, y correlacionaron las unidades ASTA con la fracción roja y amarilla de las mismas muestras determinadas mediante HPLC y espectrofotometría. Las unidades ASTA y la cantidad total de carotenoides totales están directamente relacionadas. Sin embargo, cuando se analiza el cociente entre la fracción roja y la amarilla (R/Y) no se encuentra relación con las unidades ASTA, pero con ambas medidas se puede determinar las pérdidas de color durante el proceso de almacenaje, secado, etc. Gómez Ladrón de Guevara y Pardo-González [241], estudiaron y compararon los parámetros del sistema CIE-Yxy, Y, x, y, así como L^* , a^* y b^* del espacio CIELab, ambos propuestos por la CIE. Con la representación de las coordenadas en el espacio CIELab lograron identificar las distintas fases de maduración de las 13 variedades estudiadas.

En cuanto a las pérdidas de color durante el almacenamiento, autores como Siggie et al. [242] o Ramana Rao et al. [236] han estudiado éstas pérdidas en función

de la temperatura, una variable crucial en el proceso de almacenamiento. Otros autores como Khan et al. [243] han establecido la relación entre los parámetros del espacio de color CIELAB y la cantidad de carbono orgánico, CaCO_3 y minerales en la composición del suelo, siendo posible la caracterización de las diferentes capas de del suelo empleadas para compost. González-Dugo et al. [244] estudiaron el efecto del riego en el color del pimiento para pimentón en las etapas de pre-cosecha.

Si nos referimos al efecto y cambios de color en productos procesados, donde uno de los principales ingredientes es el pimentón, encontramos que Fernández-López et al. [245] analizaron en términos de coordenadas del espacio CIELAB la influencia del pimentón en chorizos elaborados en España durante la etapa de curado. Por otra parte, aunque también en términos de parámetros L^* , a^* y b^* , Park et al. [246] han estudiado los cambios en el contenido de carotenoides, ácido ascórbico y cualidades de los frutos del género *Capsicum annuum* L. cultivados en Corea, que se han sometido a un proceso de escabechado. Mouly et al. [18] caracterizaron extractos de pimentón presentes en zumos de naranja, adulterados para conseguir un color más rojo-naranja, mediante HPLC y estudiando el perfil de los carotenoides. La escala de estándares que se usó para comparar colores fue la propuesta por el Departamento de Agricultura de los Estados Unidos (USDA) [48]. Hayashi et al. [247] también analizaron el color producido por la adición de pimentón a distintas muestras de alimentos como tortas de arroz, preparado para refrescos gaseosos edulcorados o pimientos escabechados, mediante HPLC. En éste caso, los valores de color se midieron usando acetona y las medidas de color definidas en los Métodos Oficiales para Aditivos Alimentarios [248].

I.8. Bibliografía

1. J. Salas-Salvadó, P.García Lorda, J. M. Sánchez Ripollés. *La Alimentación a través de la Historia*. Glosa, S.L. Barcelona, 2005.
2. F. López de Gomara. *Historia General de las Indias*. Linkua, S.L. Barcelona, 2008.
3. Inca Garcilaso de la Vega. *Comentarios Reales*, volumen I. Biblioteca de Ayacucho, 1º Edición. Caracas, 1976.
4. J. de Acosta. *Historia Natural de las Indias*. Madrid, 1608.
5. Real Decreto 26-09-1984, num. 2242/1984. Presidencia del Gobierno (BOE 22-12-1984, num. 306, P. 36997).
6. A. Green. *El Libro de las Especies: Hierbas Aromáticas y Especies*. Robinbook, S.L. Barcelona, 2007.
7. M. Iglesias. *Especies & Hierbas Aromáticas*. Lea, S.A. Buenos Aires, 2012.
8. F. Nuez Viñals, R. Gil Ortega, J. Costa García. *El Cultivo de Pimientos Chiles y Ajíes*. Aedos, S.A. Barcelona, 1996.
9. J. G. París. *Intercambio y Difusión de Plantas de Consumo entre el Nuevo y el Viejo Mundo*. Servicio de Extensión Agraria, Ministerio de Agricultura, Pesca y Alimentación, 1991.
10. B. Pickersgill. Review of Andrés Somos 'The Paprika'. *Experimental Agriculture*, **1985**, 21, 411-413.
11. G. Menéndez Pidal. *Hacia una Nueva Imagen del Mundo*. Real Academia de la Historia. Milenio, S.L. Madrid, 2003.
12. B. C. Kang, S.H. Nahm, J. H. Hug, H. S. Yoo, J. W. Yu, M. H. Lee, B. D. Kim. An interspecific (*Capsicum annuum* × *Capsicum chinese*) F2 linkage map in pepper using RFLP and AFLP markers. *Theoretical and Applied Genetics*, **2001**, 102, 531-539.
13. P. W. Bosland. Chiles: A diverse crop. *Horticultural Technology*, **1992**, 2, 6-10.
14. P. W. Bosland, D. DeWitt. *Peppers of the World: An Identification Guide*. Ten Speed Press, California, 1996.
15. www.FAO.org
16. D. R. Decoteau, M. J. Kasperbauer, P. G. Hunt. Bell pepper plant development over mulches of diverse colors. *Scientia Horticultura*, **1990**, 25, 460-462.

17. M. Weissenberg, I. Schaeffler, E. Menagem, M. Barzilai, A. Levy. Isocratic non-aqueous reversed-phase high-performance liquid chromatographic separation of capsanthin and capsorubin in red peppers (*Capsicum annuum* L.), paprika and oleoresin. *Journal of Chromatography A*, **1997**, 757, 89-95.
18. P. P. Mouly, E. M. Gaydou, J. Corsetti. Characterization of paprika (*Capsicum annuum*) extract in orange juices by liquid chromatography of carotenoid profiles. *Journal of Agricultural and Food Chemistry*, **1999**, 47, 968-976.
19. J. M. Guadayol, J. Caixach, J. Ribe, J. Cabañas, J. Rivera. Extraction, separation and identification of volatile organic compounds from paprika oleoresin (Spanish type). *Journal of Agricultural and Food Chemistry*, **1997**, 45, 1868-1872.
20. H. Nishino, M. Murakoshi, H. Tokuda, Y. Satomi. Cancer prevention by carotenoids. *Archives of Biochemistry and Biophysics*, **2009**, 483, 165-168.
21. C.A. Reilly. Quantitative analysis of capsaicinoids in fresh peppers, oleoresin *Capsicum* and pepper spray products. *Journal of Forensic Science*, **2001**, 46, 502-509.
22. S. T. Akar, A. Gorgulo, T. Akar, S. Celik. Decolorization of reactive Blue 49 contaminated solutions by *Capsicum annuum* seeds: Batch and continuous mode biosorption applications. *Chemical Engineering Journal*, **2011**, 168, 125-133
23. Reglamento (CEE) núm. 2081/1992, 14-07-1992 (DOCE 24-07-1992, L.208, P. 0001 – 0008).
24. Reglamento (CE) núm. 510/2006, 20-03-2006 (DOCE 31-03-2006, L.93, P.0012-0025).
25. BOE 05-01-2006, núm 4, P. 778, Ministerio de Agricultura, Pesca y Alimentación.
26. Reglamento (CE) núm. 982/2007 21-08-2007. (DOCE 22-08-2007, L. 217, P. 0022-0023).
27. Reglamento (CE) núm. 464/2001 07-03-2001. (DOCE 8-03-2001, L. 66, P. 0029-0030).
28. Orden Ministerial 08-05-1987 Ministerio de Agricultura, Alimentación y Medio Ambiente. (BOE 23-05-1987, núm. 123, pág. 15139).
29. Reglamento (CE) núm. 1485/2004 20-08-2004. (DOCE 21-08-2004, L. 273, P. 0007-0008).

30. Reglamento (CE) núm. 2400/1996 17-12-1996. (DOCE 18-12-1996, L. 327, P. 0011).
31. Reglamento (CE) núm. 417/2006 10-03-2006. (DOCE 11-03-2006, L. 72, P. 0008-0009).
32. BOE 20-08-2009, núm. 201, sec. III. P. 71663, Ministerio de Medio Ambiente y Medio Rural y Marino.
33. BOE 05-06-2008, núm. 136. P.26058, Ministerio de Agricultura, Pesca y Alimentación.
34. M. Zapata, S. Bañón, P. Cabrera. *El Pimiento para el Pimentón*. Mundiprensa, Madrid, 1992.
35. J. V. Serradilla. *El Pimentón de La Vera*. Colección Monográfica. Junta de Extremadura, 1998.
36. Ministerio de Agricultura, Alimentación y Medio Ambiente (www.magrama.gob.es), Febrero, 2013.
37. M. C. Lomas, J. C. Del Pueyo, S. Sanz. El secado del pimiento en la Rioja. De la tradición a la industrialización. *Alimentación y Equipamiento Tecnológico*, **1995**, XIV, 55-59.
38. M. I. Mínguez-Mosquera, M. Jaren-Galán, M. Garrido-Fernández, D. Hornero-Mendez. *Carotenoides en el Pimentón. Factores Responsables de su Degradación*. CSIC. Madrid, 1996.
39. J. P. Fernández-Trujillo, D. Escarabajal. El proceso tradicional de elaboración del pimentón de Murcia y sus posibles innovaciones. *Grasas y Aceites*, **2006**, 57, 433-442.
40. C. Vindras, N. Sinoir. *Tools to Integrate Organoleptic Quality Criteria in Breeding Programs*. Technical Boocklets no.1, Institut Technique de l'Agriculture Biologique, Paris, 2011.
41. J. J. Rackis, D. J. Sessa, D. H. Honig. Flavor problems of vegetable food proteins. *Journal of the American Oil Chemists' Society*, **1979**, 56, 262-271.
42. P. A. Luning, T. Ebbenhorst-Seller, T. de Rijk, J. P. Roozen. Effect of hot-air drying on flavour compounds of bell peppers (*Capsicum annuum*). *Journal of the Science of Food and Agriculture*, **1995**, 68, 355-365.

43. R. L. Jarret, L. Irvin J., T. L. Potter, S. C. Cermak. Seed oil and fatty acid composition in *Capsicum* spp. *Journal of Food Composition and Analysis*, **2013**, 30, 102–108.
44. N. P. Bekker, N. T. Ul'Chenko, A. I. Glushenkova. Physicochemmmical properties and composition of lipids from *Capsicum annuum* seeds. *Journal of Chemistry of Natural Compounds*, **2001**, 37, 131-133.
45. R. G. Buttery, R. M. Seifert, D. G. Guadagni, L. C. Ling. Characterization of some volatile constituents of bell peppers. *Journal of Agriculture and Food Chemistry*, **1969**, 17, 1322-1327.
46. H. E. Nursten. Volatile Compounds in Food. Qualitative Data by S. van Straten, H. Maarse, J. C. de Beauveser, C. A. Visscher. Division for Nutrition and Food Research TNO, Zeist, Holanda, 1983.
47. Institute of Medicine. *Dietary Reference Intakes*. The National Academies Press. Washington, D.C., 2005. (www.nap.edu)
48. Department of Agriculture, USA, (www.ars.usda.gov/nutrientdata), Agricultural Research Service, 2008.
49. D. B. Rodriguez-Amaya. International Life Sciences Institute OMNI project. *Guide to Carotenoid Analysis in Foods*. ILSI Press, Washington D.C., 2001.
50. F. Márkus, H. G. Daood, J. Kapitány, P. A. Biacs. Change in the carotenoid and antioxidant content of spice red pepper (Paprika) as a function of ripening and some technological factors. *Journal of Agricultural and Food Chemistry*, **1999**, 47, 100-107.
51. K. M. Phillips, D. M. Ruggio, M. Ashraf-Khorassani, D. B. Haytowitz. Difference in folate content of green and red sweet peppers (*Capsicum annuum*) determined by liquid chromatography-mass spectrometry. *Journal of Agricultural and Food Chemistry*, **2006**, 54, 9998-10002.
52. H. G. Daood, M. Vinkler, F. Markus, E. A. Hebshi, P.A. Biacs. Antioxidant vitamin content of spice red pepper (paprika) as affected by technological and varietal factors. *Food Chemistry*, **1996**, 55, 365-372.
53. J. A. Osuna-García, M. M. Wall, C. A. Waddell. Endogenous levels of tocopherols and ascorbic acid during fruit ripening of new Mexican-Type chile

- (*Capsicum annuum* L.) cultivars. *Journal of Agricultural and Food Chemistry*, **1998**, 46, 5093-5096.
54. P. A. Biacs, B. Czinkotai, A. Hoschke. Factors affecting stability of colored substances in paprika powders. *Journal of Agricultural and Food Chemistry*, **1992**, 40, 363-367.
55. H. G. Daood, P. A. Biacs, N. Kiss-Kutz, F. Hajdú, B. Czinkotai. Lipid and antioxidant content of red pepper. *Biological Role of Plant Lipids*, **1989**, 491-494.
56. K. Iwai, T. Suzuki, H. Fujiwake. Formation and accumulation of pungent principle of hot pepper fruits, capsaicin and its analogues, in *Capsicum annuum* var. *annuum* cv. Karayatsubusa at different growth stages after flowering. *Agricultural and Biological Chemistry*, **1979**, 43, 2493-2498.
57. J.D. Batchelor, B.T. Jones. Determination of the Scoville heat value for hot sauces and chilies: an HPLC experiment. *Journal of Chemical Education*, **2000**, 77, 266-267.
58. G. W. Witney, P. J. Hoffman, B. N. Wolstenholme. Mineral distribution in avocado trees with reference to calcium cycling and fruit quality. *Scientia Horticulturae*, **1990**, 44, 279-291.
59. H. B. Tukey, Jr. The leaching of substances from plants. *Annual Review of Plant Physiology*, **1970**, 21, 305-324.
60. B. L. Boese, H. L. Ii, D. T. Specht, R. C. Randall, M.H. Winsor. Comparison of aqueous and solid-phase uptake for hexachlorobenzene in the tellinid clam *Macoma nasuta* (conrad): A mass balance approach. *Environmental Toxicology and Chemistry*, **1990**, 9, 221-231.
61. B. Estrada, F. Pomar, J. Díaz, F. Merino and M. A. Bernal. Effects of mineral fertilizer supplementation on fruit development and pungency in 'Padrón' peppers. *The Journal of Horticultural Science & Biotechnology*, **1998**, 73, 493-497.
62. P. Kirschbaum-Titze, C. Hiepler, E. Mueller-Seitz, M. Petz. Pungency in Paprika (*Capsicum annuum*). 1. Decrease of capsaicinoid content following cellular disruption. *Journal of Agricultural and Food Chemistry*, **2002**, 50, 1260-1263.
63. E. S. Poyrazoğlu, O. Yemi, Ç. Kadakal, N. Artık. Determination of capsaicinoid profile of different chilli peppers grown in Turkey. *Journal of the Science of Food and Agriculture*, **2005**, 85, 1435-1438.

64. M. Contreras-Padilla, E. M. Yahia. Changes in capsaicinoids during development, maturation and senescence of chile peppers and relation with peroxidase activity. *Journal of Agricultural and Food Chemistry*, **1998**, 46, 2075-2079.
65. F. Korel, N. Bağdatlıoğlu, M. Ö. Balaban, Y. Hişil. Ground red peppers: capsaicinoids content, Scoville scores and discrimination by an electronic nose. *Journal of Agricultural and Food Chemistry*, **2002**, 50, 3257-3261.
66. K. Iwai, T. Suzuki, H. Fujiwake. Simultaneous microdetermination of capsaicin and its four analogues by using high-performance liquid chromatography and gas chromatography—mass spectrometry. *Journal of Chromatography A*, **1979**, 172, 303-311.
67. V. K. Attuquayefio, K. A. Buckle. Rapid sample preparation method for HPLC analysis of capsaicinoids in *Capsicum* fruits and oleoresins. *Journal of Agricultural and Food Chemistry*, **1987**, 35, 777-779.
68. I. Perucka, W. Oleszek. Extraction and determination of capsaicinoids in fruit of hot pepper *Capsicum annuum* L. by spectrophotometry and high-performance liquid chromatography. *Food Chemistry*, **2000**, 71, 287-291.
69. V. Illés, M. H. Gnayfeed, B. Mészáros, G. Horváth, P. A. Biacs. Extraction of pungent spice paprika by supercritical carbon dioxide and subcritical propane. *The Journal of Supercritical fluids*, **2002**, 23, 143-152.
70. R. Karnkca, M. Rayanakorn, S. Watanesk, Y. Vaneesorn. Optimization of High-Performance Liquid Chromatographic parameters for the determination of capsaicinoid compounds using the simplex method. *Analytical Sciences*, **2002**, 18, 661-665.
71. J. Juangsamoot, C. Ruangviriyachai, S. Techawongstien, S. Chanthai. Determination of capsaicin and dihydrocapsaicin in some hot chilli varieties by RP-HPLC-PDA after magnetic stirring extraction and clean up with C₁₈ cartridge. *Journal of International Food Research*, **2012**, 19, 1217-1226.
72. R. I. Santamara, M. D. Reyes-Duarte, E. Barzana, D. Fernando, F. M. Gama, M. Mota, A. López-Mungua. Selective enzyme-mediated extraction of capsaicinoids and carotenoids from chili Guajillo Puya (*Capsicum annuum* L.) using ethanol as solvent. *Journal of Agricultural and Food Chemistry*, **2000**, 48, 3063-3067.

73. G. F. Barbero, M. Palma, C. G. Barroso. Determination of capsaicinoids in peppers by microwave-assisted extraction–high-performance liquid chromatography with fluorescence detection. *Analytica Chimica Acta*, **2006**, 578, 227-233.
74. D. E. Breithaupt, W. Schwack Determination of free and bound carotenoids in paprika (*Capsicum annuum* L.) by LC/MS. *European Food Research and Technology*, **2000**, 211, 52-55.
75. B. Cámara, R. Monéger. Free and esterified carotenoids in green and red fruits of *Capsicum annuum*. *Phytochemistry*, **1978**, 17, 91-93.
76. P. G. Hoffman, M. C. Lego, W. G. Galetto. Separation and quantitation of red pepper major heat principles by reverse-phase high-pressure liquid chromatography. *Journal of Agricultural and Food Chemistry*, **1983**, 31, 1326-13.
77. M. D. Collins, L. M. Wasmund, P. W. Bosland. Improved method for quantifying capsaicinoids in *Capsicum* using high performance liquid chromatography. *Horticultural Science*, **1995**, 30, 137-139.
78. Z. A. Al Othman, Y. B. Hadj Ahmed, M. Abdelaty Habila, A. A. Ghafar. Determination of capsaicin and dihydrocapsaicin in *Capsicum* fruit samples using high performance liquid chromatography. *Molecules*, **2011**, 16, 8919-8929.
79. Official Analytical Methods of the American Spice Trade Association, ASTA. www.astaspice.org/. Julio 2012.
80. B. H. Davies, S. Matthews, J. T. O. Kirk. The nature and biosynthesis of the carotenoids of different colour varieties of *Capsicum annuum*. *Phytochemistry*, **1970**, 9, 797-805.
81. U. Schweiggert, R. Carle, A. Schieber. Characterization of major and minor capsaicinoids and related compounds in chili pods (*Capsicum frutescens* L.) by high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Analytica Chimica Acta*, **2006**, 557, 236-244.
82. T. A. Betts. Pungency quantitation of hot pepper sauces using HPLC. *Journal of Chemical Education*, **1999**, 76, 240-244.
83. M. Peusch, E. Miller-Seitz, M. Petz. Determination of capsaicinoids in low pungency paprika. *Zeitschrift für Lebensmittel Untersuchung und Forschung*, **1996**, 202, 334-336.

84. K. S. Rymal, R. D. Cosper, D. A. Smith. Injection-extraction procedure for rapid determination of relative pungency in fresh Jalapeno peppers. *Journal Association Official Analytical Chemists*, **1984**, 67, 658-659.
85. A. Garcés-Claver, M. S. Arnedo-Andrés, J. Abadía, R. Gil-Ortega, A. Álvarez-Fernández. Determination of capsaicin and dihydrocapsaicin in *Capsicum* fruits by liquid chromatography–electrospray/time-of-flight mass spectrometry. *Journal of Agriculture and Food Chemistry*, **2006**, 54, 9303–9311.
86. F. Mendel, L. Kap-Rang Lee, K. Hyun-Jeong, L. In-Seon, N. Kozukue. Anticarcinogenic effects of glycoalkaloids from potatoes against human cervical, liver, lymphoma, and stomach cancer cells. *Journal of Agriculture and Food Chemistry*, **2005**, 53, 6162–6169.
87. S. Singh, R. Jarret, V. Russo, G. Majetich, J. Shimkus, R. Bushway, B. Perkin. Determination of capsinoids by HPLC-DAD in *Capsicum* species. *Journal of Agricultural and Food Chemistry*, **2009**, 57, 3452-345.
88. N. Wu, J. Dempsey, P. M. Yehl, A. Dovletoglou, D. Ellison, J. Wyvratt. Practical aspects of fast HPLC separations for pharmaceutical process development using monolithic columns. *Analytica Chimica Acta*, **2004**, 523, 149-156.
89. D. J. Hart, K. J. Scott. Development and evaluation of an HPLC method for the analysis of carotenoids in foods, and the measurement of the carotenoid content of vegetables and fruits commonly consumed in the UK. *Food Chemistry*, **1995**, 54, 101-111.
90. A. Topuz, C. Dincer, K. S. Özdemir, H. Feng, M. Kushad. Influence of different drying methods on carotenoids and capsaicinoids of paprika (Cv., Jalapeno). *Food Chemistry*, **2011**, 129, 860-865.
91. M. I. Minguez-Mosquera, D. Hornero-Mendez. Separation and quantification of the carotenoid pigments in red peppers (*Capsicum annuum* L.), paprika, and oleoresin by reversed-phase HPLC. *Journal of Agricultural and Food Chemistry*, **1993**, 41, 1616-1620.
92. C. Fisher, J. A. Kocis. Separation of paprika pigments by HPLC. *Journal of Agricultural and Food Chemistry*, **1987**, 35, 55-57.
93. G. Mózsik, F. Morón, M. Fiegler, T. Jávora, T. L. Nagy, I. Páthy, F. Tárnok. Membrane-bound ATP-dependent energy systems and gastric cytoprotection by

prostacyclin, atropine and cimetidine in the rat. *International Journal of Tissue Reactions*, **1983**, 5, 263-278.

94. D. Giuffrida, P. Dugo, G. Torre, C. Birgnaridi, A. Cavazzo, C. Corradini, G. Dugo. Characterization of 12 *Capsicum* varieties by evaluation of their carotenoid profile and pungency determination. *Food Chemistry*, **2013**, 140, (4), 794-802.

95. F. A. Müller-Stock, R. K. Joshi, J. Büchi. Studies on the ingredients of *Capsicum* by quantitative gas chromatographic determination of individual homologous and analogous of capsaicinoid mixtures of natural origin and vanillylamids as falsification. *Journal of Chromatography A*, **1971**, 63, 281-287.

96. A. M. Krajewska, J. J. Powers. Pentafluorobenzoylation of capsaicinoids for gas chromatography with electron-capture detection. *Journal of Chromatography A*, **1988**, 457, 279-286.

97. B. V. Thomas, A. A. Schreiber, C. P. Weisskopf. Simple method for quantitation of capsaicinoids in peppers using capillary gas chromatography. *Journal of Agricultural and Food Chemistry*, **1998**, 46, 2655-2663.

98. D. R. Cremer, K. Eichner. Formation of volatile compounds during heating of spice paprika (*Capsicum annum*) powder. *Journal of Agricultural and Food Chemistry*, **2000**, 48, 2454-2460.

99. J. M. Guadayol, J. Caixach, J. Ribé, J. Cabañas, J. Rivera. Extraction, separation and identification of volatile organic compounds from paprika oleoresin (Spanish type). *Journal of Agricultural and Food Chemistry*, **1997**, 45, 1868-1872.

100. M. J. Oruña-Concha, J. López-Hernández, J. A. Simal-Lozano, J. Simal-Gándara, M. J. González-Castro, C. de la Cruz- García. Determination of Volatile Components in Fresh, Frozen, and Freeze-Dried Padrón-Type Peppers by Gas Chromatography-Mass Spectrometry Using Dynamic Headspace Sampling and Microwave Desorption. *Journal of Chromatographic Science*, **1998**, 36, 583-588.

101. J. Pino, E. Sauri-Duch, R. Marbot. Changes in volatile compounds of Habanero chile pepper (*Capsicum chinense* Jack. cv. Habanero) at two ripening stages. *Food Chemistry*, **2006**, 94, 394-398.

102. J. Pino, M. González, L. Ceballos, A. R. Centurión-Yah, J. Trujillo-Aguirre, L. Latournerie-Moreno, E. Sauri-Duch. Characterization of total capsaicinoids, colour

and volatile compounds of Habanero chilli pepper (*Capsicum chinense* J.) cultivars grown in Yucatan. *Food Chemistry*, **2007**, 104, 1682-1686.

103. R. G. Buttery, R. M. Seifert, D. G. Guadagni, L. C. Ling. Characterization of some volatile constituents of bell peppers. *Journal of Agricultural and Food Chemistry*, **1969**, 17, 1322-1327.

104. N. Kocsis, M. Amtmann, Z. Mednyánszky, K. Korány. GC-MS investigation of the aroma compounds of Hungarian red paprika (*Capsicum annuum*) cultivars. *Journal of Food Composition and Analysis*, **2002**, 15, 195-203.

105. K. M. Kang, H. H. Baek. Aroma quality assessment of Korean fermented red pepper paste (gochujang) by aroma extract dilution analysis and headspace solid-phase microextraction–gas chromatography–olfactometry. *Food Chemistry*, **2014**, 145, 488-495.

106. M. Ziino, C. Conduro, V. Romeo, G. Tripodi, A. Verzera. Volatile compounds and capsaicinoid content of fresh hot peppers (*Capsicum annuum* L.) of different Calabrian varieties. *Journal of the Science of Food and Agriculture*, **2009**, 89, 774-780.

107. G. Heltai, K. Percsich. Moderated pressure microwave digestion system for preparation of biological samples. *Talanta*, **1994**, 41, 1067-1072.

108. C. Baluja-Santos, A. González-Portal. Application of hydride generation to atomic-absorption spectrometric analysis of wines and beverages: A review. *Talanta*, **1999**, 39, 329-339.

109. I. Kojima, T. Uchida, C. Iida. Pressurized microwave digestion of biological samples for metal determination. *Analytical Sciences*, **1988**, 4, 211-214.

110. A. Abu-Samra, J. S. Morris, S. R. Koirtyohann. Wet ashing of some biological samples in a microwave oven. *Analytical Chemistry*, **1975**, 47, 1475-1477.

111. P. Barret, L. J. Davidowski Jr., K. W. Penaro, K. W. Penaro, T. R. Copeland. Microwave oven-based wet digestion technique. *Analytical Chemistry*, **1978**, 50, 1021-1023.

112. I. Aydin. Comparison of dry, wet and microwave digestion procedures for the determination of chemical elements in wool samples in Turkey using ICP-OES technique. *Microchemical Journal*, **2008**, 90, 82-87.

113. H. Altundag, M. Tuzen. Comparison of dry, wet and microwave digestion methods for the multi element determination in some dried fruit samples by ICP-OES. *Food and Chemical Toxicology*, **2011**, 49, 2800–2807.
114. V. Filgueiras, J.L. Capelo, I. Lavilla, C. Bendicho. Comparison of ultrasound-assisted extraction and microwave-assisted digestion for determination of magnesium, manganese and zinc in plant samples by flame atomic absorption spectrometry. *Talanta*, **2000**, 53, 433-441.
115. I. Lavilla, A. V. Filgueiras, C. Bendicho. Comparison of digestion methods for determination of trace and minor metals in plant samples. *Journal of Agricultural and Food Chemistry*, **1999**, 47, 5072-5077.
116. G. F. Antonious, T. S. Kochhar. Mobility of heavy metals from soil into hot pepper fruits: A field study. *Bulletin of Environmental Contamination and Toxicology*, **2009**, 82, 59-63.
117. W. J. de Melo, P. de Stéfani Aguiar, G. M. Peruca de Melo, V. Peruca de Melo. Nickel in a tropical soil treated with sewage sludge and cropped with maize in long-term field study. *Soil Biology and Biochemistry*, **2007**, 39, 1341-1347.
118. J. Khairiah, M. K. Zalifah, Y. H. Yin, A. Aminah. The uptake of heavy metals by fruit type vegetables grown in selected agricultural areas. *Pakistan Journal of Biological Sciences*, **2004**, 7, 1438-1442.
119. U. Divrikli, N. Horzum, M. Soylak, L. Elci1. Trace heavy metal contents of some spices and herbal plants from western Anatolia, Turkey. *International Journal of Food Science & Technology*, **2006**, 41, 712–716.
120. M. Soylak, Z. Cihan, E. Yilmaz. Evaluation of trace element contents of some herbal plants and spices retailed in Kayseri, Turkey. *Environmental Monitoring and Assessment*, **2012**, 184, 3455-3461.
121. J. P. Diaz-Alarcon, M. Navarro-Alarcon, H. Lopez-Garcia de la Serrana, M. C. Lopez-Martinez. Determination of Selenium levels in vegetables and fruits by hydride generation atomic absorption spectrometry. *Journal of Agricultural and Food Chemistry*, **1994**, 42, 2848-2851.
122. I. L. Sziklai, M. Ördögh, E. Molnár, E. Szabó Distribution of trace and minor elements in Hungarian spice paprika plants. *Journal of Radioanalytical and Nuclear Chemistry*, **1988**, 122, 233-238.

123. R. Chizzola, H. Michitsch, C. Franz. Monitoring of metallic micronutrients and heavy metals in herbs, spices and medicinal plants from Austria. *European Food Research and Technology*, **2003**, 216, 407-411.
124. L. Liu, X. Chen, L. Juli, X. Deng, W. Duan, S. Tan. Determination of capsaicin and dihydrocapsaicin in *Capsicum annum* and related products by capillary electrophoresis with a mixed surfactant system. *Food Chemistry*, **2010**, 119, 1228-1232.
125. A. Laskaridou-Monnerville. Determination of capsaicin and dihydrocapsaicin by micellar electrokinetic capillary chromatography and its application to various species of *Capsicum*, Solanaceae. *Journal of Chromatography A*, **1999**, 838, 293-302.
126. K. L. Bajaj, G. Kaup. Colorimetric determination of capsaicin in *Capsicum* fruits with de Folin-Ciocalteu reagent. *Microchimica Acta [Wien]*, **1979**, 81-86.
127. H. A. A. Gibbs, L. W. O'Garro. Capsaicin content of West Indies hot pepper cultivars using colorimetric and chromatographic techniques. *Horticultural Science*, **2004**, 39, 132-135.
128. S. C. S. Tsou, T. L. Hong, T. Berke. Use of near infrared reflectance to measure capsaicinoids in pepper (*Capsicum* spp). *Capsicum and Eggplant Newsletter*, **1997**, 16, 56-59.
129. X. Wang, L. Xue, X. He, M. Liu Vitamin C content estimation of chilies using Vis/NIR spectroscopy. *International Conference on Electric Information and Control Engineering (ICEICE)*, **2011**, 1894-1897.
130. T. Ignat Z. Schmilovitch, J. Fefoldia, B. Steiner, S. Alkalai-Tuviac. Non-destructive measurement of ascorbic acid content in bell peppers by VIS-NIR and SWIR spectrometry. *Postharvest Biology and Technology*, **2012**, 74, 91-99.
131. R. Baranski, M. Baranska, I. Schulz. Changes in carotenoid content and distribution in living plant tissue can be observed and mapped in situ using NIR-FT-Raman spectroscopy. *Planta*, **2005**, 222, 448-457.
132. I. Schulz, R. Baranski, M. Baranska. Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis. *Biopolymers*, **2005**, 77, 212-221.
133. J. R. Pigottt. *Statistical Procedures in Food Research*, Elsevier Applied Science, London, **1986**.

134. A. Kauffman. Multivariate Statistics as a classification tool in the food laboratory. *Journal of AOAC International*, **1997**, 80, 665-675.
135. B. R. Kowalski, C. F. Bender. Pattern Recognition. A powerful approach to interpreting chemical data. *Journal of the American Chemical Society*, **1972**, 94, 5632-5639.
136. B. R. Kowalski, C. F. Bender. Pattern Recognition II. A powerful approach to interpreting chemical data. *Journal of the American Chemical Society*, **1972**, 95, 686-693.
137. W. Luo, S. Huan, H. Fu, G. Wen, H. Cheng, J. Zhou, H. Wu, G. Shen, R. Yu. Preliminary study on the application of near infrared spectroscopy and pattern recognition methods to classify different types of apple samples. *Food Chemistry*, **2011**, 128, 555-561.
138. J. J. Mangas, R. Rodríguez, B. Suárez, A. Pincinelli, E. Dapena. Study of the phenolic profile of cider Apple cultivars at maturity by multivariate techniques. *Journal of Agricultural and Food Chemistry*, **1999**, 47, 4046-4052.
139. D. Ariana, D. E. Guyer, B. Shrestha. Integrating multispectral reflectance and fluorescence imaging for defect detection on apples. *Computers and Electronics in Agriculture*, **2006**, 50, 148-161.
140. Y. Zheng, G. Ruan, B. Li, C. Xiong, S. Chen, M. Luo, Y. Li, F. Du. Multicomposition analysis and pattern recognition of Chinese geographical indication product: vinegar. *European Food Research and Technology*, **2014**, 238, 337-344.
141. M. J. Benito, M. C. Ortiz, M. S. Sánchez, L. A. Sarabia, M. Iñiguez. Typification of vinegars from Jerez and Rioja using classical chemometric techniques and neural network methods. *Analyst*, **1999**, 124, 547-552.
142. R. Natera, R. Castro, M. V. García-Moreno, M. J. Hernández, C. García-Barroso. Chemometric studies of vinegars from different raw materials and processes of production. *Journal of Agricultural and Food Chemistry*, **2003**, 51, 3345-3351.
143. M. C. García- Parrila, G. A. González, F. J. Heredia, A. M. Troncoso. Differentiation of wine vinegars based on phenolic composition. *Journal of Agricultural and Food Chemistry*, **1997**, 45, 3487-3492.

144. L. Giansante, D. Di Viceno, G. Bianchi. Classification of monovarietal Italian olive oils by unsupervised (PCA) and supervised (LDA) chemometrics. *Journal of the Science and Food*, **2003**, 83, 905-911.
145. R. M. Alonso-Salces, K. Héberger, M. V. Holland, J. M. Moreno-Rojas, C. Mariani, G. Bellan, F. Reniero, C. Guillou. Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chemistry*, **2010**, 118, 956-965.
146. F. Peña, S. Cárdenas, M. Gallego, M. Varcárcel. Characterization of olive oil classes using a ChemSensor and pattern recognition techniques. *Journal of the American Oil Chemists' Society*, **2002**, 79, 1103-1108.
147. F. Marini, F. Baslestrieri, R. Bucci, A. L. Magrì, D. Marini. Supervised pattern recognition to discriminate the geographical origin of rice bran oils: a first study. *Microchemical Journal*, **2003**, 74, 239-248.
148. M. Decker, B. Ronn, S. S. Jorgensen. Thermally assisted in-line methylation and gas chromatography with statistical data analysis for determination of fatty acid distribution and fingerprint of plant seeds and oils. *European Food Research and Technology*, **2000**, 211, 366-373.
149. M. V. Baroni, C. Arrua, M. L. Nores, P. Fayé, M. P. Díaz, G. A. Chiabrande, D. A. Wanderlin. Composition of honey from Córdoba (Argentina): Assessment of North/South provenance by chemometrics. *Food Chemistry*, **2009**, 114, 727-733.
150. Y. Yücel, P. Sultanoğlu. Characterization of honeys from Hatay Region by their physicochemical properties combined with chemometrics. *Food Bioscience*, **2013**, 1, 16-25.
151. Y. Yücel, P. Sultanoğlu. Characterization of Hatay honeys according to their multi-element analysis using ICP-OES combined with chemometrics. *Food Chemistry*, **2013**, 140, 231-237.
152. M. Chudzinska, D. Baralkiewicz. Estimation of honey authenticity by multielements characteristics using inductively coupled plasma-mass spectrometry (ICP-MS) combined with chemometrics. *Food and Chemical Toxicology*, **2010**, 48, 284-290.
153. E. F. Boffo, L. A. Tavares, A. C. T. Tobias, M. M. C. Ferreira, A. G. Ferreira. Identification of components of Brazilian honey by ¹H NMR and classification of its

- botanical origin by chemometric methods. *Food Science and Technology*, **2012**, 49, 55-63.
154. L. F. Huang, M. J. Wu, K. J. Zhong, X. J. Sun, Y. Z. Liang, Y. H. Dai, K. L. Huang, F. Q. Guo. Fingerprint developing of coffee flavor by gas chromatography–mass spectrometry and combined chemometrics methods. *Analytica Chimica Acta*, **2007**, 588, 216-223.
155. J. S. Ribeiro, F. Augusto, T. J. G. Salva, M. M. C. Ferreira. Prediction models for Arabica coffee beverage quality based on aroma analyses and chemometric. *Talanta*, **2012**, 101, 253-260.
156. A. P. Lelis Rodrigues de Oliveira, P. Cesar Correa, E. Lázaro-Reis, G. H. Horta de Oliveira. Comparative Study of the Physical and Chemical Characteristics of Coffee and Sensorial Analysis by Principal Components. *Food Analytical Methods*, **2015**, 8, 1303-1314.
157. R. Muñiz-Valencia, J. M. Jurado, S. G. Ceballos-Magaña, A. Alcázar, J. Hernandez-Díaz. Characterization of Mexican coffee according to their mineral contents by means of multilayer perceptrons artificial neural networks. *Journal of Food Composition and Analysis*, **2014**, 34, 7-11.
158. R. Briandet, E. K. Kemsley, R. H. Wilson. Approaches to adulteration detection in instant coffees using infrared spectroscopy and chemometrics. *Journal of the Science of Food and Agriculture*, **1996**, 71, 359-366.
159. J. L. Valentin, R. J. Walting. Provenance establishment of coffee using solution ICP-MS and ICP-AES. *Food Chemistry*, **2013**, 141, 98-104.
160. D. L. McKay, J. B. Blumberg. The role of tea in human health: an update. *Journal of the American College of Nutrition*, **2002**, 21, 1-13.
161. P. Valera, F. Pablos, A. G. González. Classification of tea samples by their chemical composition using discriminant analysis. *Talanta*, **1996**, 43, 415-419.
162. A. Alcázar, O. Ballesteros, J. M. Jurado, F. Pablos, M. J. Martín, J. L. Vilches, A. Navalón. Differentiation of Green, White, Black, Oolong, and Pu-erh Teas According to Their Free Amino Acids Content. *Journal of Agricultural and Food Chemistry*, **2007**, 55, 5960–5965.

163. J. S. McKenzie, J. M. Jurado, F. Pablos. Characterisation of tea leaves according to their total mineral content by means of probabilistic neural networks. *Food Chemistry*, **2010**, 121, 859-864.
164. A. Palacios-Morillo, A. Alcázar, F. Pablos, J. M. Jurado. Differentiation of tea varieties using UV–Vis spectra and pattern recognition techniques. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2013, 103, 79-83.
165. M. Suhaj, M. Koreňovská. Application of elemental analysis for determination of wine origin. *Acta Alimentaria*, **2005**, 34, 393-401.
166. M. J. Latorre, C. García Jares, B. Medina, C. Herrero. Pattern Recognition analysis applied to classification of wines from Galicia (Northwestern Spain) with certified brand of origin. *Journal of Agricultural and Food Chemistry*, **1994**, 42, 1451-1455.
167. S. Frías, J. E. Conde, J. J. Rodríguez-Bencomo, F. García-Montelongo, J. P. Trujillo. Classification of commercial wines from the Canary Islands (Spain) by chemometric techniques using metallic contents. *Talanta*, **2003**, 59, 335-344.
168. A. González, A. Llorens, M. L. Cervera, A. Armenta, M. de la Guardia. Elemental fingerprint of wines from the protected designation of origin Valencia. *Food Chemistry*, **2009**, 112, 26-34.
169. J. Šperkova, M. Suchánek. Multivariate classification of wines from different Bohemian regions (Czech Republic). *Food Chemistry*, **2005**, 93, 659–663.
170. J. M. Jurado, A. Alcázar, A. Palacios- Morillo, F. de Pablos. Classification of Spanish DO white wines according to their elemental profile by means of support vector machines. *Food Chemistry*. **2012**, 135, 898-903.
171. M. Urbano, M. Luque de Castro, P. M. Pérez, J. García-Olmo, M. A. Gómez-Nieto. Ultraviolet–visible spectroscopy and pattern recognition methods for differentiation and classification of wines. *Food Chemistry*, **2006**, 97, 166-175.
172. D. Cazzolino, H. E. Smyth, M. Gishen. Feasibility study on the use of visible and Near-Infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *Journal of Agricultural and Food Chemistry*, **2003**, 51, 7703-7708.

173. K. Fernandez, E. Agosin. Quantitative analysis of red wine tannins using Fourier-Transform Mid-Infrared spectrometry. *Journal of Agricultural and Food Chemistry*, **2007**, 55, 7294-7300.
174. J. M. Jurado, O. Ballesteros, A. Alcázar, F. Pablos, M. J. Martín, J. L. Vílchez, A. Navalón. Differentiation of certified brands of origins of Spanish white wines by HS-SPME-GC and chemometrics. *Analytical and Bioanalytical Chemistry*, **2008**, 390, 961-970.
175. A. de Villiers, P. Majek, F. Lynen, A. Crouch, H. Lauer, P. Sandra. Classification of South African red and white wines according to grape variety based on the non-coloured phenolic content. *European Food Research and Technology*, **2005**, 221, 520-528.
176. A. Alcázar, J. M. Jurado, A. Palacios-Morillo, F. Pablos, M.J. Martín. Differentiation of blonde beers according to chemical quality indicators by means of pattern recognition techniques. *Food Analytical Methods*, **2012**, 5, 795-799.
177. A. Alcázar, J. M. Jurado, A. Palacios-Morillo, F. Pablos, M. J. Martín. Recognition of the geographical origin of beer based on support vector machines applied to chemical descriptors. *Food control*, **2012**, 23, 258-262.
178. A. Biancolillo, R. Bucci, A. L. Magrì, A. D. Magrì, F. Marini. Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication. *Analytica Chimica Acta*, **2014**, 820, 23-31.
179. S. G. Ceballos-Magaña, F. Pablos, J. M. Jurado, M. J. Martín, A. Alcázar. Characterization of tequila according to their mayor volatile composition using multilayer perceptron neural networks. *Food Chemistry*, **2013**, 136, 1309-1315.
180. U. Contreras, A. O. Barbosa-García, J. L. Pichardo-Molina, G. Ramos-Ortiz, J. L. Maldonado, M. A. Meneses-Nava. N. E. Ornelas-Soto, P. L. Lopez de Alba. Screening method for identification of adulterate and fake tequilas by using UV–VIS spectroscopy and chemometrics. *Food Research International*, **2010**, 43, 2356–2362.
181. J. M. Jurado, O. Ballesteros, A. Alcázar, F. Pablos, M. J. Martín, J. L. Vílchez, A. Navalón. Characterization of aniseed-flavoured spirit drinks by headspace solid-phase microextraction gas chromatography–mass spectrometry and chemometrics. *Talanta*, **2007**, 72, 506-511.

182. D. Yucesoy, B. Ozen. Authentication of a Turkish traditional aniseed flavoured distilled spirit, raki. *Food Chemistry*, **2013**, 141, 1461-1465.
183. M. J. C. Pontes, S. R. B. Santos, M. C. U. Araújo, L. F. Almeida, R. A. C. Lima, E. N. Gaião, U. T. C. P. Souto. Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. *Food Research International*, **2006**, 39, 182-189.
184. J. C. Moore, J. Sink, M. Lipp. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *Journal of Food Science*, **2012**, 77, R118-R126.
185. M. A. de la Fuente, M. Juárez. Authenticity assessment of dairy products. *Critical Reviews in Food Science and Nutrition*, **2005**, 45, 563-586.
186. P. M. Santos, E. R. Pereira-Filho, L. E. Rodriguez-Saona. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. *Food Chemistry*, **2013**, 138, 19-24.
187. M. Bassbasi, S. Platikanov, R. Tauler, A. Oussama. FTIR-ATR determination of solid non fat (SNF) in raw milk using PLS and SVM chemometric methods. *Food Chemistry*, **2014**, 146, 250-254.
188. L. L. Botros, J. Jablonski, C. Chang, M. M. Bergana, P. Wehling, J. M. Harnly, G. Downey, P. Harrington, A. R. Potts, J. C. Moore. Exploring authentic skim and nonfat dry milk powder variance for the development of nontargeted adulterant detection methods using near-infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry*, **2013**, 61, 9810-9818.
189. G. Downey, V. Fouratier, J. D. Kelly. Detection of honey adulteration by addition of fructose and glucose using near infrared transreflectance spectroscopy. *Journal of Near Infrared Spectroscopy*, **2003**, 11, 447-456.
190. J. D Kelly, G. Downey, V. Fouratier. Initial study of honey adulteration by sugar solutions using mid-infrared (MIR) spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry*, **2004**, 52, 33-39.
191. C. Cordella, J. S. L. T. Militao, M. C. Clement, P. Drajnudel, D. Cabrol-Bass. Detection and quantification of honey adulteration via direct incorporation of sugar syrups or bee-feeding: preliminary study using high-performance anion exchange

- chromatography with pulsed amperometric detection (HPAEC-PAD) and chemometrics. *Analytica Chimica Acta*, **2005**, 531, 239-248.
192. B. López, M. J. Latorre, M. I. Fernández, M. A. García, S. García, C. Herrero. Chemometrics classification of honeys according to their type based on quality control data. *Food Chemistry*, **1996**, 55, 281-287.
193. X. Zhu, S. Li, Y. Shan, Z. Zhang, G. Li, D. Su, F. Liu. Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics. *Journal of Food Engineering*, **2010**, 101, 92-97.
194. S. A. Ordoudi, M. M Pascual, M. Z. Tsimidou. On the quality control of traded saffron by means of transmission Fourier-transform mid-infrared (FT-MIR) spectroscopy and chemometrics. *Food Chemistry*, **2014**, 150, 414-421.
195. V. A. Parthasarathy, B. Chempakam, T. John Zachariah. *Chemistry of Spices*. CABI, 2008.
196. F. Korel, N. Bagdatlioglu, M. O. Balaban, Y. Hisil. Ground Red Pepper: Capsaicinoids content, Scoville scores, and discrimination by an electronic nose. *Journal of Agricultural and Food Chemistry*, **2002**, 50, 3257-3261.
197. S. Yudthavorasit, K. Wongravee, N. Leepipatpiboon. Characteristic fingerprint based on gingerol derivative analysis for discrimination of ginger (*Zingiber officinale*) according to geographical origin using HPLC-DAD combined with chemometrics. *Food Chemistry*, 2014, 158, 101-111.
198. C. Karadas, D. Kara. Chemometric approach to evaluate trace metal concentrations in some spices and herbs. *Food Chemistry*, **2012**, 130, 196-202.
199. A. A. D'Archivio, A. Giannitto, A. Incani, S. Nisi. Analysis of the mineral composition of Italian saffron by ICP-MS and classification of geographical origin. *Food Chemistry*, **2014**, 157, 485-489.
200. E. Anastasaki, C. Kanakis, C. Pappas, L. Maggi, C. P. del Campo, M. Carmona, G. L. Alonso, M. G. Polissiou. Geographical differentiation of saffron by GC-MS/FID and chemometrics. *European Food Research and Technology*, **2009**, 229, 899-905.
201. E. Anastasaki, C. Kanakis, C. Pappas, L. Maggi, C. P. del Campo, M. Carmona, G. L. Alonso, M. G. Polissiou. Differentiation of saffron from four

- countries by mid-infrared spectroscopy and multivariate analysis. *European Food Research and Technology*, **2010**, 230, 571-577.
202. T. G. H. James. *La pintura Egipcia*. Akal S.A., Madrid, 1999.
203. J. W. von Goethe, C. Lock Eastlake. *Goethe's Theory of Colours: Translated from the German*. Murray, Baviera, 1840.
204. J. Schanda. *Colorimetry: Understanding the CIE System*, John Wiley & Sons, 2007.
205. J. Cat. *Maxwell, Sutton, and the Birth of Color Photography: A Binocular Study*. Palgrave MacMillan, Nueva York, 2013.
206. B. Berlin, P. Kay. *Basic Colors Terms: Their Universality and Evolution*. University of California Press. Berkley, 1969.
207. J. L. Calvano. *Sistemas de Orden del Color*. Universidad de Buenos Aires, Argentina, 1995.
208. T. Young. On the Theory of Light and Colors. *Philosophical Transactions of the Royal Society*, **1802**, 92, 12-48.
209. H. Hertz. Las Ondas Electromagnéticas. Volumen II, *Clásicos de las Ciencias*. Universidad Autónoma de Barcelona, 1990
210. K. Nassau. *Color for Science, Art and Technology*. Elsevier, 1997.
211. J. Gilabert. *Medida del Color*. Universidad Politécnica de Valencia, 2002.
212. J. Cohen. *Visual Color and Color Mixture: The Fundamental Color Space*. University of Illinois Press, 2001.
213. R. G. Kuehni. *Color Vision & Technology*. AATCC, 2008.
214. UNE 72.015-82. Observador colorimétrico UNE 4 (campo mayor de 4°)
215. M. I. Alfonso Galipienso, M. A. Cazorla Quevedo, O. Colomina Pardo. *Inteligencia Artificial: Modelos, Técnicas y Áreas de Aplicación*. Paraninfo, 2003.
216. A. Hicethier. *Le Cube des Couleurs*. Dessain et Tolra, Paris, 1985.
217. H. Küppers. *Atlas de Colores*, Blume, Barcelona, 1994.
218. C. E. Foss, D. Nickerson, W. C. Granville. Analysis of the Ostwald color system. *Journal of the Optical Society of America*, **1944**, 34, 361-381.
219. *Munsell Book of Color*. Munsell Color Co. Baltimore, Maryland, 1943.
220. J. M. Artigas, P. Capilla, A. Felipe, J. Pujol, *Óptica Fisiológica. Psicofísica de la Visión*. Interamericana Mac Graw-Hill, 1995

221. K. Ritcher. Cube-Root Color Spaces and chromatic adaptation. *Color Research & Application*, **1980**, 5, 25-43.
222. J. L. Saunderson, B. I. Milner. Modified chromatic value color space. *Journal of Optical Society*, **1946**, 36, 36-41.
223. OIV. Determination of Chromation Characteristics aAccording to CIELab (Method OIV-MA_AS2-11). Compendium of International Analysis of Methods-OIV, 2006.
224. CIE Standard Colorimetric observers <http://www.cis.rit.edu/mcsl/online/cie.php>
225. R. E. C. Wildman, R. Wildman, T. C. Wallace. *Handbook of Nutraceuticals and Functional Foods*. CRC Boca Raton, Florida, 1999.
226. A. Marín, F. Ferreres, F. A. Tomas, M. Gil, Characterization and quantitation of antioxidant constituents of sweet pepper (*Capsicum annuum* L.). *Journal of Agricultural and Food Chemistry*, **2004**, 52, 3861-3869.
227. H. Matsufuji, H. Nakamura, M. Chino, M. Takeda. Antioxidant activity of capsanthin and the fatty acid esters in paprika (*Capsicum annuum*). *Journal of Agricultural and Food Chemistry*, **1998**, 46, 3468-3472.
228. K. Suna, H. Youl, J. Park. Characteristics of pigment composition and colour value by the difference of harvesting times in Korean red pepper varieties (*Capsicum annuum*, L.). *Journal of European Food Research and Tehcnology*, **2008**, 43, 915-920.
229. H. Tuorila-Ollikainen, S. Mahlamäki-Kultanen, R. Kurkela. Relative importance of Color, fruity flavor and sweetness in the overall liking of soft drinks. *Journal of Food Science*, **1984**, 49, 1598-1600.
230. F. M. Clydesdale. Color as a factor in food choice. *Critical Reviews in Food Science and Nutrition*, **1993**, 33, 83-101.
231. A. González Benanvente-García, J. Fernández Hernández, G. Pérez Cañavate, J. I. Llorens Pascual de Riquelme. La horticultura en la región de Murcia. *Agrícola Vergel*, **1990**, 9, 344-348.
232. J. T. Vanderslice, D. J. Higgs, J. M. Hayes, G. Block. Ascorbic acid and dehydroascorbic acid content of foods-as-eaten. *Journal of Food Composition and Analysis*, **1990**, 3, 105-118.

233. N. I. Krinsky, E. J. Johnson. Carotenoid actions and their relation to health and disease. *Molecular Aspects of Medicine*, **2005**, 26, 459-516.
234. M. I. Mínguez-Mosquera, A. Pérez-Gálvez. Color quality in paprika oleoresins. *Journal of Agricultural and Food Chemistry*, **1998**, 46, 5124-5127.
235. J. Jang, M. Rahman, A.Y. Ko, A. M. Abd El-Aty, J. H. Park, S. K. Cho, J. H. Shim. A matrix sensitive gas chromatography method for the analysis of pymetrozine in red pepper: Application to dissipation pattern and PHRL. *Food Chemistry*, **2014**, 146, 448-454.
236. T. V. Ramana Rao, N. B. Gol, K. K. Shah. Effect of postharvest treatments and storage temperatures on the quality and shelf life of sweet pepper (*Capsicum annum* L.). *Scientia Horticulturae*, **2005**, 123, 18-26.
237. M. I. Mínguez Mosquera, M. Jaren-Galan, J. Garrido-Fernández. Competition between the Processes of Biosynthesis and Degradation of Carotenoids during the drying of peppers. *Journal of Agricultural and Food Chemistry*, **1994**, 42, 645-648
238. M. I. Mínguez Mosquera, M. I. Cols. Paprika (*Capsicum annum*). Ratio between the red and yellow carotenoids pigments. *Grasas y Aceites*, **1984**, 35, 4-10.
239. S. D. Lannes, F. L. Finger, A. R. Schuelter, V. W. D. Casa. Growth and quality of Brazilian accessions of *Capsicum chinense* fruits. *Scientia Horticulturae*, **2007**, 112, 266-270.
240. M. Krajayklang, A. Klieber, P. R. Dry. Colour at harvest and post-harvest behavior influence paprika and chilli spice quality. *Postharvest Biology and Technology*, **2000**, 20, 269-278.
241. R. Gómez Ladrón de Guevara, J. E. Pardo-González. Evolution of color during ripening of selected varieties of paprika pepper (*Capsicum annum* L.). *Journal of Agricultural and Food Chemistry*, **1996**, 44, 2049-2052.
242. G. O. Sigge, C. F. Hansmann, E. Joubert. Effect of the storage conditions, packaging material and metabisulphite treatment on the color of the dehydrated green bell peppers (*Capsicum annum* L.). *Journal of Food Quality*, **2001**, 24, 205-218.
243. M. A. I. Khan, K. Ueno, S. Horimoto, F. Komai. CIELAB color variables as indicators of compost stability. *Waste Management*, **2009**, 29, 2969-2975.

244. M. González-Dugo, F. Orgaz, E. Fereres. Respones of pepper to deficit irrigation for paprika production *Scientia Horticulturae*, **2007**, 114, 77-82.
245. J. Fernández-López, J. A. Pérez Álvarez, E. Sayas-Barberá, F. López-Santoveña. Effect of paprika (*Capsicum annum*) on color of Spanish-type sausages during the resting stage. *Journal of Food Science*, **2002**, 67, 2410-2414.
246. H. Park, S. Kim, B. Moon. Changes in Carotenoids, Ascorbic Acids, and Quality Characteristics by the Pickling of Paprika (*Capsicum annum* L.) Cultivated in Korea. *Journal of Food Science*, **2011**, 76, 1075-1080.
247. T. Hayashi, K. Hayashi, J. Fujita, M. Ono, H. Oka. An HPLC Method for the analysis of paprika color in food using capsanthin as an indicator. *Journal Liquid Chromatography and Related Technologies*, **2001**, 24, 2347-2361.
248. R. Wood, L. Foster, A. Damant, P. Key. *Analytical Methods for Food Additives*. Woodhead Publishing Ltd, Cambridge, 2004.

II. Quimiometría

II.1. Introducción a la quimiometría

La automatización y computarización de los laboratorios ha llevado consigo algunas consecuencias. Una de ellas es la rápida adquisición de gran cantidad de datos analíticos sobre muchos componentes de una misma muestra. Cuadras [1] define el análisis multivariante como la parte de la estadística y el análisis de datos que estudia, analiza, representa e interpreta los datos que resultan de observar más de una variable estadística sobre una muestra de individuos. Hablamos, por tanto, de datos multivariantes cuando se dispone de varias variables medidas para una misma muestra. La adquisición de estos datos requiere de un diseño apropiado de los experimentos y procedimientos a realizar, para poder extraer la máxima información posible. En el ámbito de la química, la Quimiometría es la disciplina que tiene esta finalidad ya que se encarga de desarrollar las herramientas matemáticas necesarias [2].

Las raíces de la Quimiometría se remontan a 1969 cuando Jurs, Kowalski e Isenhour publicaron una serie de artículos en la revista *Analytical Chemistry* sobre la aplicación de máquinas de aprendizaje lineales para clasificar espectros de masas de baja resolución [3]. En ese momento se introdujo una nueva forma de pensar con la que se pretendía transformar una gran cantidad de datos en información útil. Fue en 1977 cuando Wold publicó un artículo donde utilizaba un algoritmo matemático para el análisis multivariante de datos [4]. Estas nuevas técnicas, que no pasaron inadvertidas para los investigadores, sirvieron para que los especialistas en el campo de la química analítica concluyeran que algunas teorías matemáticas, como el reconocimiento de patrones [5] o la teoría de la información [6] tenían especial relevancia en química analítica para la evaluación, optimización, selección, clasificación y decisión de procedimientos a aplicar.

Las técnicas de análisis multivariante, aunque proporcionan una información valiosa, son relativamente complejas y requieren de ciertos conocimientos previos sobre la matriz de datos a tratar ya que no todas son aplicables a cualquier conjunto de

variables. Hair et al. [7] proponen el esquema que se muestra en la figura II.1. para obtener los resultados más óptimos a la hora de trabajar con de análisis multivariante.

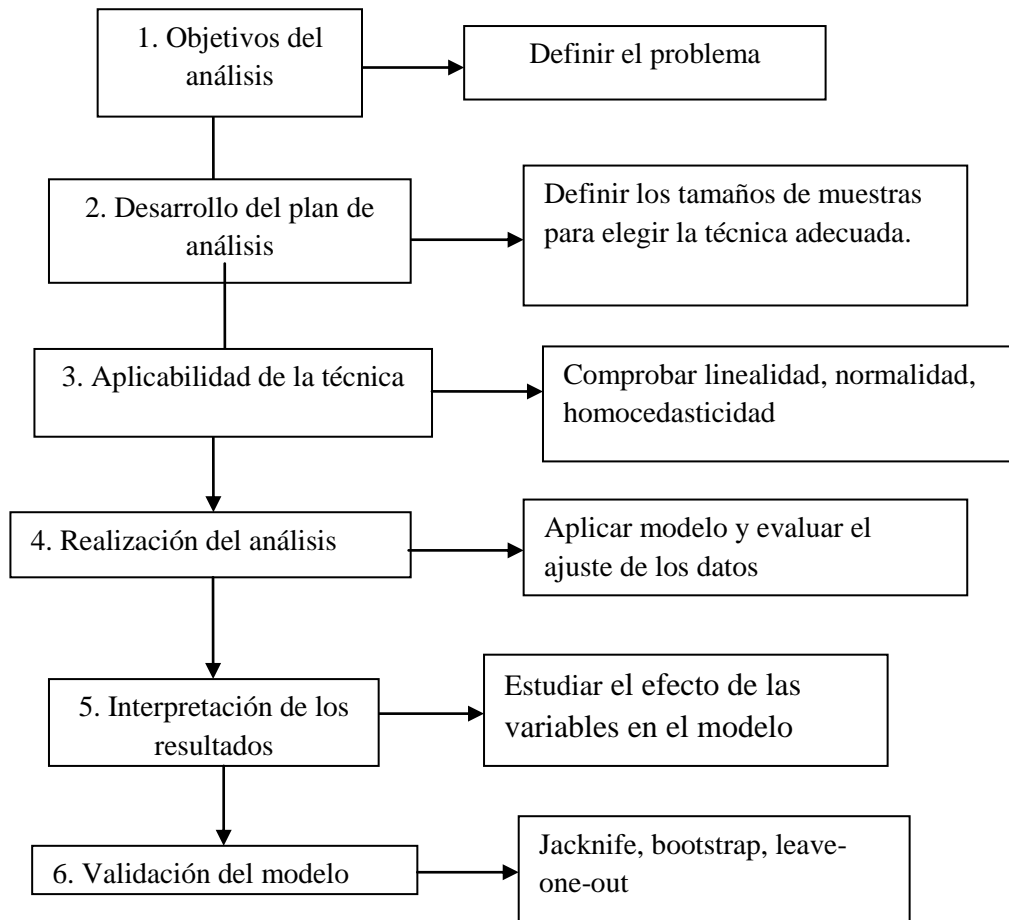


Figura II.1. Esquema de la secuencia previa a la aplicación del análisis multivariante.

En primer lugar hay que definir, en términos conceptuales, cuál es el problema y qué técnicas se pretenden usar y qué variables se van a observar. A continuación se determina el tamaño de la muestra y se evalúa la aplicabilidad de la técnica, por ejemplo, si los datos a tratar muestran una distribución normal, linealidad u homocedasticidad. En cuarto lugar, se aplica el modelo y se observa si las variables elegidas son aptas para la construcción de este y si el nivel de bondad de ajuste es el adecuado. Se interpretan los resultados, y en ocasiones hay que volver a redefinir los apartados 3 y 4, porque las variables elegidas no proporcionen los resultados esperados. Por último, una vez que el modelo se ha construido y se ha decidido que se adapta bien a los requerimientos propuestos, es necesario validarlo para, de alguna manera, asegurar que esos datos son

extensibles al resto de la población. La validación cruzada mediante jackknife o bootstrap son técnicas de validación frecuentes.

Tal y como se señala en el esquema de la figura II.1., la normalidad es una característica de los datos que es necesario determinar porque en función de si la distribución es normal o no, se aplica estadística básica previa, paramétrica o no paramétrica, respectivamente. Para verificar que las variables observadas siguen de forma conjunta una distribución normal multivariante, es posible comprobar la normalidad de las variables de forma individual ya que si una de las variables analizadas individualmente no presenta distribución normal, el conjunto tampoco lo hará [8]. Por tanto, la normalidad univariante es condición necesaria pero no suficiente para que el conjunto multivariante sea normal. Los contrastes univariantes son más sencillos de aplicar que los contrastes de significación multivariante [9], que son bastante más complejos. El hecho de poder analizar la normalidad univariante introduce una ventaja importante en el análisis estadístico.

II.2. Pruebas de normalidad

Como ya se ha mencionado, en función de la distribución que presenten los datos a tratar habrá que realizar una estadística, previa a los métodos de reconocimiento de patrones, paramétrica o no paramétrica. Las pruebas comúnmente más empleadas son las siguientes:

- *Contrastes de asimetría y curtosis*

Antes de empezar a aplicar test estadísticos, es posible y de manera gráfica, tener cierta idea de si la distribución será normal o no, aunque finalmente haya que verificar con pruebas estadísticas. Este tipo de contrastes permiten determinar si la forma de la distribución de las observaciones muestrales se aleja significativamente de la de un modelo normal en lo que a su simetría y curtosis se refiere. Se tiene una distribución simétrica cuando los valores que están a la misma distancia de la media tienen igual frecuencia; en caso contrario se trata de una curva asimétrica positiva o negativa. La curtosis analiza el grado de concentración que presentan los valores alrededor de la zona

central de la distribución. Se suele considerar el coeficiente de curtosis relativo, para poder comparar la forma de la distribución con la de la distribución normal (campana de Gauss). El coeficiente de curtosis se calcula según:

$$Curtosis = g_2 = \frac{\sum_{i=1}^N (x - \bar{x})^4}{Ns^4} - 3 \quad (2.1)$$

donde s es la desviación estándar típica, N el número total de muestras y \bar{x} el valor medio de las muestras. En función de los resultados obtenidos se definen tres tipos de distribuciones [10]:

- Si $g_2 = 0$ entonces se trata de una distribución mesocúrtica que presenta el mismo grado medio de concentración alrededor de los valores centrales de la variable, que en el caso de una distribución normal.
- Si $g_2 > 0$ entonces se trata de una distribución leptocúrtica que presenta un mayor grado de concentración alrededor de los valores centrales de la variable.
- Si $g_2 < 0$ entonces se trata de una distribución platicúrtica que presenta un reducido grado de concentración alrededor de los valores centrales de la variable.

En la figura II.2. se muestran distintas distribuciones, normales y no normales con asimetrías y curtosis de distintos tipos.

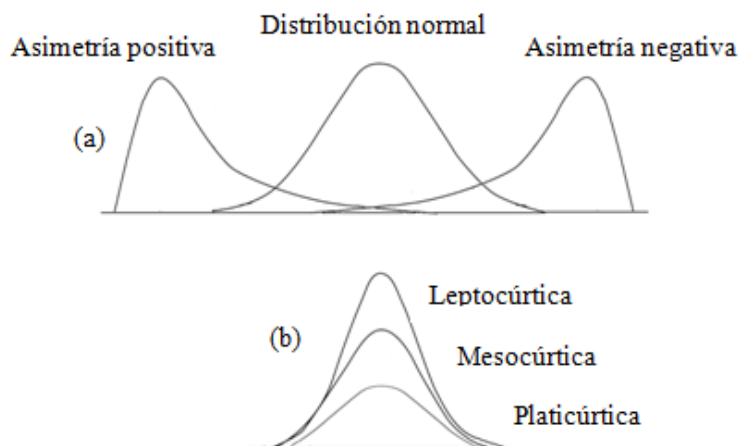


Figura II.2. (a) Distribuciones simétricas y asimétricas; (b) distribuciones con distintos coeficientes de curtosis.

▪ *Contraste de Shapiro-Wilks*

El test de Shapiro-Wilks fue publicado en 1965 por Samuel Shapiro y Martin Wilks y es considerado uno de los test más potentes para el contraste de normalidad [11]. Para efectuar la prueba es necesario calcular la media y la varianza muestral, S^2 , y ordenar las observaciones en orden creciente. Se calculan las diferencias entre: el primer y el último valor; el segundo y el penúltimo y así sucesivamente y se corrigen con unos coeficientes tabulados por Shapiro-Wilk. El estadístico de prueba es:

$$W = \frac{D}{nS^2} \quad (2.2)$$

donde D es la suma de las diferencias corregidas. La hipótesis nula se rechazará cuando se obtengan valores de W menores que el valor crítico tabulado para un tamaño muestral y nivel de significación, α , dados, concluyendo que la muestra no proviene de una distribución normal. Este contraste es el más adecuado cuando el tamaño de muestra es pequeño, no superior a 50.

▪ *Prueba de Anderson-Darling*

La prueba de Anderson-Darling se usa para probar si un conjunto de datos proviene de una población con una distribución específica (por ejemplo una población normal) [12]. Se basa en la comparación de la distribución de probabilidades empírica con la distribución de probabilidades teórica. Presenta la ventaja de una buena sensibilidad y la desventaja de que los valores críticos deben ser calculados para cada tipo de distribución. En el caso de la distribución normal, estos valores críticos están tabulados. El estadístico de la prueba es, en este caso:

$$A^2 = -N - S \quad (2.3)$$

siendo,

$$S = \sum_{i=1}^N \frac{(2i-1)}{N} [(\ln(F(Y_i)) + \ln(1 - F(Y_{N+1-i})))] \quad (2.4)$$

donde F es la función distribución de la distribución específica, evaluada en el i -ésimo elemento en orden ascendente de la muestra, ya que Y_i se refiere a los datos ordenados crecientemente. La hipótesis nula, que supone que la muestra aleatoria proviene de una distribución específica, se rechaza cuando el estadístico A^2 toma un valor menor que el valor crítico tabulado para un tamaño muestral y nivel de significación, α , dados.

▪ *Prueba de Lilliefors*

Es una modificación del contraste de bondad de ajuste de Kolmogorov-Smirnov, propuesta por Hubert Lilliefors, profesor de estadística en la Universidad George Washington [13]. Se emplea para probar la hipótesis nula de que los datos provienen de una población distribuida normalmente, cuando la hipótesis nula no especifica los parámetros de la distribución normal, es decir, no especifica el valor esperado y la varianza de la distribución. Este contraste compara la función de distribución empírica muestral con la teórica de una población normal, de manera que se rechazaría la hipótesis nula de normalidad si el valor experimental del estadístico, L , es mayor que el teórico tabulado para un número de observaciones y nivel α de confianza dado, $L_{\text{tab}}(\alpha, \nu)$ (tabla A2.1. del anexo 2). Este contraste resulta muy apropiado cuando el tamaño de muestra es grande.

El estadístico, L , se calcula según:

$$L_n = \sup |F_n(x) - \hat{F}(x)| \quad (2.5)$$

donde \hat{F} es la distribución normal cuyas media y varianza son las estimadas, es decir con Z normal típica.

Existen más pruebas para determinar la distribución de las muestras, como el contraste de Kolmogorov-Smirnov o la prueba Cramer-von Mises, entre otras.

II.3. Contrastes de significación paramétricos

II.3.1. Contraste t de Student

Las distribuciones t de Student se atribuyen al estadístico William Sealy Gosset, nacido en Canterbury y conocido por su sobrenombre Student, el cual solía emplear para publicar en la revista *Biometrika* [14].

La idea de la distribución t de Student surge ante el problema de necesitar una distribución que pudiera ser usada cuando el tamaño de la muestra fuera pequeño y la varianza desconocida y, además de poder estimarse a partir de los datos [15].

Algunas de las propiedades que se pueden destacar de la distribución t son [16]:

- a) La distribución t de Student tiene un tamaño diferente para cada tamaño de muestra.
- b) La distribución t de Student tiene igual forma que la distribución normal (campana simétrica) pero cuando el tamaño de muestra es grande, refleja una mayor variabilidad. A medida que el tamaño de la muestra, n, se hace más grande, la distribución t de Student se acerca a la distribución normal.
- c) La distribución t de Student tiene una media de $t = 0$ y una desviación estándar mayor que uno.

Un contraste t de Student se aplica cuando se quiere comparar la media de una muestra con un valor supuesto verdadero. La hipótesis nula es que la media de la población de la que se extrae la muestra es igual al valor verdadero μ . Se calcula el estadístico t como:

$$t = \frac{|\bar{x} - \mu| \sqrt{n}}{s} \quad (2.6)$$

Y se compara con el valor tabulado $t(\alpha, n-1)$ (tabla A2.2 del anexo 2). Si el valor calculado es mayor que el tabulado se rechaza la hipótesis nula.

En el caso de la comparación de los valores promedio de dos muestras, el cálculo del estadístico t, depende de varios factores muestrales, como la igualdad o no de la varianza o los tamaños muestrales. Para decidir si se asume o no que las varianzas son iguales, hay que aplicar la prueba F de Snedecor [17], dando lugar a dos posibilidades:

1) Si las varianzas no difieren se calcula una varianza promedio

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (2.7)$$

Se calcula el estadístico t:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2.8)$$

Este valor de t se compara con el valor tabulado t para $\alpha=0,05$ y $n_1 + n_2 - 2$ grados de libertad (tabla A2.2. del anexo 2). Si $t < t_{\text{crit}}$ se acepta la hipótesis nula, es decir, las dos medias son estadísticamente iguales.

2) Si las varianzas difieren se calcula el estadístico t como:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (2.9)$$

Ahora los grados de libertad efectivos de la varianza combinada se deben calcular mediante la ecuación de Welch-Satterthwaite [18]:

$$v_{\text{eff}} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}\right)} \quad (2.10)$$

El valor de t se compara con $t(\alpha, v_{\text{eff}})$ (tabla A2.2 del anexo 2).

II.3.2. Análisis de la varianza

El análisis de la varianza, ANOVA, es una herramienta de estadística paramétrica empleada para comparar a la vez varias medias muestrales. Se trata de comprobar si todas las medias son iguales o si existe, al menos, alguna que introduzca un sesgo significativo. Cuando se tiene un único factor, se trata de una generalización de la prueba t de *Student* para un diseño con más de dos muestras.

Aunque el objetivo del ANOVA es comparar medias y no varianzas, cuando se aplica ANOVA al conjunto de datos, se descompone la varianza total de un conjunto de datos en dos contribuciones, una debida al error aleatorio y que se mide a través de la varianza de repetitividad (INTRA) y otra debida a un error sistemático provocados por factores controlados (INTER). Se comparan dichas varianzas mediante un ensayo F de Snedecor-Fisher para comprobar si el factor controlado incrementa la varianza por un sesgo distinto a los errores aleatorios. Si la igualdad de las varianzas no se acepta, es necesario averiguar mediante ensayos de comparación múltiple, qué nivel de variación del factor está introduciendo dicho error [19].

II.3.2.1. ANOVA de un factor

La prueba de ANOVA de un solo factor nos permite comparar las medias de h grupos ($h > 2$). Para poder aplicar ANOVA hay que asegurarse de que las muestras se distribuyen normalmente, lo que se debería comprobar aplicando algún test de normalidad como el de Lilliefors. Otro requisito es que las varianzas se distribuyan homogéneamente, lo que se puede comprobar mediante la prueba C de Cochran (sección II.5.3). Si estas condiciones se cumplen, la media obtenida para cada grupo será:

$$\bar{x}_i = \frac{\sum_j^n x_{ij}}{n} \quad (2.11)$$

La media global obtenida para todos los grupos se calculará como:

$$\bar{x} = \frac{\sum_i^h \bar{x}_i}{h} \quad (2.12)$$

Para cada grupo se tiene una varianza calculada como:

$$s_i^2 = \frac{\sum_j^n (x_{ij} - \bar{x}_i)^2}{n-1} \quad (2.13)$$

En el ANOVA de un factor la hipótesis nula es que las medias de los h grupos se extraen de una población de media μ y varianza σ^2 . Esta varianza se puede estimar de dos formas, estudiando la variación dentro de las muestras y la variación entre muestras [20].

La varianza debida al error puramente aleatorio, s_{PE}^2 , también conocida como varianza dentro de las muestras (o del grupo), se calcula como promedio de las varianzas individuales s_i^2

$$s_{PE}^2 = \frac{\sum_i^h s_i^2}{h} = \frac{\sum_i^h \sum_j^n (x_{ij} - \bar{x}_i)^2}{h(n-1)} \quad (2.14)$$

La varianza debida al factor, conocida como varianza entre muestras (o entre grupos), es una estimación de la varianza de la población, σ^2 . Si todas las medidas se extraen de una población de varianza, σ^2 , entonces sus medias (obtenidas variando un factor controlado) proceden de una población de varianza σ^2/n . La varianza debida al factor se calcula como:

$$s_{Factor}^2 = \frac{n \sum_i^h (\bar{x}_i - \bar{x})^2}{h-1} \quad (2.15)$$

Bajo la hipótesis nula, la variabilidad entre grupos no deberá superar significativamente a la variabilidad dentro del grupo. Es decir, el hecho de variar el factor no introduce un error significativo en comparación con el puramente aleatorio. Se comprueba esto mediante un test de Fisher, calculando F como:

$$F = \frac{S_{Factor}^2}{S_{PE}^2} \quad (2.16)$$

Se compara F con el valor tabulado $F(0.05, h-1, h(n-1))$ (tabla A2.3. del anexo 2) para contraste de una cola. Si el valor de F calculado es menor que el tabulado, no existen diferencias significativas entre las dos varianzas, es decir, el hecho de variar el factor no introduce un sesgo significativo.

Si se ha rechazado la hipótesis nula de igualdad de medias, es decir, que F sea mayor que el valor de F crítico, se supone que existe al menos una diferencia, pero no se sabe cuántas, ni entre qué niveles del factor. Para poder ver lo que ocurre se ordenan las medias de forma creciente y se comparan las diferencias de los valores adyacentes con una cantidad denominada la mínima diferencia significativa. Esta se calcula como:

$$MDS = S_{PE} \sqrt{\frac{2}{n}} \cdot t_{h(n-1)} \quad (2.17)$$

Otra posibilidad es aplicar la prueba de la genuina diferencia significativa de Tukey, como se explica en el apartado II.3.2.3.

II.3.2.2. ANOVA de dos factores anidados

Cuando varios grupos o muestras se asignan a diferentes combinaciones de dos factores, el ANOVA correspondiente se denomina de dos factores y se diferencia del ANOVA de un factor porque se estudia la posible interacción entre dichos factores los cuales pueden estar afectando a los resultados de los experimentos. En este caso la varianza se descompone en la varianza de cada factor, la correspondiente a la interacción de ambos y la del error aleatorio [21]. Existen multitud de modelos de ANOVA, por ejemplo en función del número de factores (ANOVA de un factor, ANOVA de dos factores...) o del muestreo de niveles, (cuántos K niveles se asignan a cada factor), entre otros. Un tipo de ANOVA es el modelo anidado en donde los factores subordinados aparecen anidados en el factor superior. El análisis ANOVA anidado de dos factores, tiene una hipótesis nula, H_0 , para cada factor; los subgrupos dentro de cada grupo tienen igual media y los grupos tienen la misma media [22]. Este

tipo de ANOVA es útil para diseñar y determinar errores en experimentos analíticos o procesos industriales, ya que informa del grado de variabilidad dentro de cada factor.

Si se realizan una serie de experimentos variando un factor A en $i=(1-a)$ niveles y en cada factor A se varía un factor B en $j=(1-b)$ niveles realizando, además $k=(1-r)$ replicados, se tiene una situación como la que se muestra en el esquema de la figura II.3. R_{ijk} sería el k -ésimo resultado (replicado) del experimento obtenido para el j -ésimo nivel del factor B en el i -ésimo nivel del factor A.

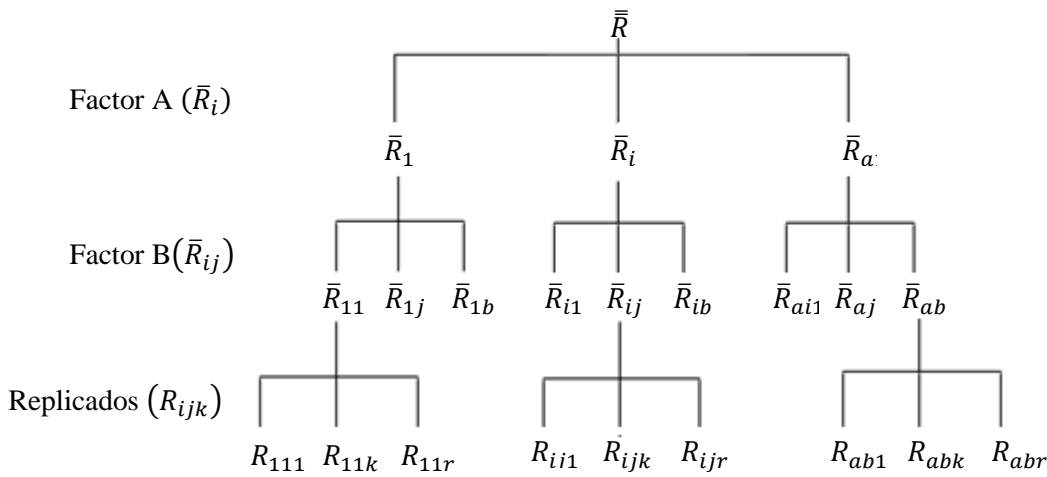


Figura II.3. Esquema anidado de un experimento variando dos factores con replicación.

Del esquema anterior se puede deducir que \bar{R}_{ij} es la media aritmética de los r replicados del nivel j del factor B en el nivel i del factor A. \bar{R}_i es la media aritmética de los resultados obtenidos en el nivel i del factor A. \bar{R} es la media total de los resultados obtenidos

$$\bar{R}_{ij} = \frac{\sum_{k=1}^r R_{ijk}}{r} \quad (2.18)$$

$$\bar{R}_i = \frac{\sum_{j=1}^b \bar{R}_{ij}}{b} \quad (2.19)$$

$$\bar{\bar{R}} = \frac{\sum_{i=1}^a \bar{R}_i}{a} \quad (2.20)$$

Como se ha dicho anteriormente, la varianza total se descompone en diferentes niveles. En el caso del ANOVA anidado de dos factores el modelo estadístico es el siguiente[23]:

$$R_{ijk} = \mu + A_i + B_{ij} + \varepsilon_{ijk} \quad (2.21)$$

siendo,

μ : la media global.

A_i : el efecto del nivel i -ésimo del factor A ($i=1 \dots i \dots a$).

B_{ij} : el efecto producido por el nivel j -ésimo del factor B en el nivel i -ésimo del factor A ($j= 1 \dots j \dots b$).

ε_{ijk} : termino de error aleatorio para la k -ésima observación del j -ésimo subnivel del nivel i .

Las ecuaciones empleadas para el cálculo de estas varianzas se resumen en la tabla II.1.

Tabla II.1. Cálculo de las varianzas para un ANOVA de dos factores anidados.

Fuente	Varianza	Cuadrado Medio Esperado
Factor A	$S_{Factor A}^2 = \frac{rb \sum_{i=1}^a (\bar{R}_i - \bar{\bar{R}})^2}{a-1}$	$rb\sigma_{Factor A}^2 + r\sigma_{Factor B}^2 + \sigma_{rep}^2$
Factor B	$S_{Factor B}^2 = \frac{r \sum_{i=1}^a \sum_{j=1}^b (\bar{R}_{ij} - \bar{R}_i)^2}{a(b-1)}$	$r\sigma_{Factor B}^2 + \sigma_{rep}^2$
Error puro	$S_{rep}^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r (R_{ijk} - \bar{R}_{ij})^2}{ab(r-1)}$	σ_{rep}^2
Total	$S_{Total}^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r (R_{ijk} - \bar{\bar{R}})^2}{abr-1}$	

Una vez realizados los cálculos es necesario comprobar mediante una prueba de significación si existe efecto significativo de los factores. Se comparan las varianzas obtenidas en cada nivel del experimento, es decir A o B, con la varianza del error puro mediante una prueba F de dos colas (tabla A2.4. del anexo 2). Para asegurar que no existe un sesgo significativo introducido por el factor, es necesario que F_{cal} sea menor que F_{tab} para un 95% de nivel de confianza.

Si el resultado de $F_{exp} > F_{cal}$, y por tanto al ir variando los niveles de un factor se introducen en los resultados errores de tipo sistemático, es necesario averiguar qué nivel del factor es el responsable de dichas diferencias. ANOVA no proporciona esta información por lo que hay recurrir a otros análisis denominados análisis *post-hoc* como Tukey-HSD, entre otros.

II.3.2.3. Prueba de la genuina diferencia significativa de Tukey

La prueba de Tukey-HSD (Host Significance Difference) [24] es una prueba de comparación múltiple que compara dos a dos las diferencias de un conjunto de grupos. En el caso del ANOVA compara niveles de un factor. Se calculan, por tanto, todas las diferencias de medias muestrales entre los niveles del factor estudiado y los niveles que sean significativamente diferentes del resto serán los responsables del sesgo.

Cuando se quiere estudiar si hay o no diferencias significativas entre dos niveles q y t del factor principal de un sistema de dos factores anidados, hay que comprobar si existen diferencias significativas entre los valores promedio de estos dos niveles, \bar{R}_t y \bar{R}_q , y demostrar que ambos provienen, o no, de una misma población de distribución normal, de media μ y varianza σ^2 . Si \bar{R}_t es la media muestral de una población de media μ_t y \bar{R}_q es la media muestral de una población de media μ_q , y no existe diferencias entre los niveles debe cumplirse la hipótesis nula de $H_0: \mu_t = \mu_q$ [25].

Se define el intervalo de confianza de Tukey como:

$$\mu_t - \mu_q \in \left(\bar{R}_t - \bar{R}_q \right) \pm \frac{q_{r,v}^{\alpha}}{\sqrt{2}} \sqrt{S^2 \left(\frac{1}{n_t} + \frac{1}{n_q} \right)} \quad (2.22)$$

donde $q_{r,v}^{\alpha}$ es el valor de la distribución del rango estandarizado [26] que depende del número de repeticiones, r , de los grados de libertad, v , utilizados para obtener la desviación estándar, s , que se usa en el cálculo de este intervalo y del nivel de confianza elegido, α .

Como la varianza empleada en la ecuación 2.22 es una estimación de la población y se debe considerar todas las muestras, se sustituye por la varianza del error puro. Además, normalmente el número de muestras de los niveles es el mismo luego $n_t = n_q = n$ y el intervalo de confianza de Tukey queda como:

$$\mu_t - \mu_q \in (\bar{R}_t - \bar{R}_q) \pm q_{r,v}^{\alpha} S_{rep} \sqrt{\left(\frac{1}{n}\right)} \quad (2.23)$$

Como $\mu_t - \mu_q = 0$, el intervalo de confianza de Tukey debe contener el cero, si no es así, se rechaza la H_0 .

$$(\bar{R}_t - \bar{R}_q) - q_{r,v}^{\alpha} S_{rep} \sqrt{\left(\frac{1}{n}\right)} < 0 < (\bar{R}_t - \bar{R}_q) + q_{r,v}^{\alpha} S_{rep} \sqrt{\left(\frac{1}{n}\right)} \quad (2.24)$$

En la práctica se parte de la $H_0: \mu_t = \mu_q$ y se calcula un valor de q experimental de acuerdo con la siguiente fórmula:

$$q_{exp} = \frac{\bar{R}_t - \bar{R}_q}{S_{rep} \sqrt{\left(\frac{1}{n}\right)}} \quad (2.25)$$

Se compara este valor experimental de q con el valor tabulado [26] (tabla A2.5. del anexo 2). Si $q_{exp} > q_{tab}$ se rechazaría H_0 , con lo que las dos medias presentarían diferencias significativas entre sí. El nivel que presente diferencias significativas con el resto, se considera como causa del efecto significativo en la varianza total.

Si se tienen tamaños muestrales distintos esta prueba pasa a denominarse prueba de Tukey-Kramery n se calcula según:

$$n_{Tukey-Kramer} = \frac{2}{\sum_{i=1}^2 \frac{1}{n_i}} \quad (2.26)$$

II.4. Contrastes de significación no paramétricos

En contraposición a la estadística paramétrica, que presupone una distribución normal de los datos, la estadística no paramétrica es aquella que no exige una distribución concreta de la población de datos. En estadística no paramétrica se emplean los conceptos mediana, en lugar de la media, percentiles y moda [27]. Las pruebas de significación que se han explicado en el apartado III.3. no son válidas para el caso de una distribución que no sea normal, por lo que es necesario describir otro tipo de pruebas. En los siguientes apartados se explicarán las pruebas no paramétricas que se han usado en este trabajo.

II.4.1. Prueba U de Mann-Whitney

La prueba U de Mann-Whitney fue propuesta por H.B. Mann y D.R. Whitney en 1947 [28] y está basada en la prueba Wilcoxon, para muestras independientes. La prueba U de Mann-Whitney permite contrastar si dos grupos independientes de datos proceden de poblaciones donde una variable aleatoria cuantitativa se distribuye idénticamente [29]. En cierta medida puede considerarse como la alternativa no paramétrica del estadístico t de Student, para la comparación de medias de dos distribuciones. Parte de una hipótesis nula, H_0 , de que las muestras provienen de poblaciones idénticas. Es necesario calcular el estadístico U tal y como sigue:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad (2.27)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \quad (2.28)$$

donde n_1 y n_2 son los tamaños muestrales y R_1 y R_2 la suma de los rangos de las muestras 1 y 2, respectivamente. Los estadísticos U_1 y U_2 se transforman en valores z , según la ecuación 2.29 [30].

$$z_i = \frac{U_i - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + (n_2 + 1))}{12}}} \quad (2.29)$$

Por último se compara el estadístico z_i mínimo entre z_1 y z_2 con el valor de z tabulado (tabla A2.6. del anexo 2) para un tamaño de muestra y nivel de significación dados.

- Si $z_{\text{cal}} < z_{\text{tab}}$ se acepta la H_0 , de que ambas muestras provienen de la misma población
- Si $z_{\text{cal}} > z_{\text{tab}}$ se rechaza la H_0 y no se puede aceptar que las dos muestras pertenezcan a la misma población.

II.4.2. Test de Kruskal- Wallis

La prueba de Kruskal-Wallis, o prueba H de Kruskal-Wallis, propuesta por William Kruskal y W. Allen Wallis [31] puede considerarse la alternativa no paramétrica al ANOVA y una extensión de la prueba U de Mann Whitney cuando los grupos a comparar son más de dos. El objetivo de la prueba es, por tanto, contrastar la diferencia de localización de K poblaciones por medio de n muestras independientes de tamaños relativos.

El procedimiento para el cálculo del estadístico H se basa en una suma de rangos, como se expresa en la siguiente ecuación [21]:

$$H = \frac{12}{N(N+1)} \left[\sum_i^{n_i} \frac{\left(\sum_j^{n_j} R_{ij} \right)^2}{n_j} \right] - 3(N+1) \quad (2.30)$$

donde

n_j es el número de casos para un grupo particular.

N es el número total de casos.

$\sum_j^{n_j} R_{ij}$ es la suma de los rangos de todas las j observaciones en el grupo i .

El estadístico H se distribuye como χ^2 para $n-1$ grados de libertad. Una vez se ha fijado el nivel de significación, α , y se ha elegido el tipo de prueba, si $H_{cal} > \chi^2_{tab}$ para un α y v dados (tabla A2.7. del anexo 2), se rechaza la hipótesis nula consistente en que los K grupos provienen de la misma población. Antes de utilizar la prueba de Kruskal-Wallis es necesario asegurarse que el tamaño de muestra sea el adecuado. Cuando se aplica Kruskal-Wallis para tres grupos, es necesario que haya al menos cinco observaciones para cada uno. Si existen más de tres grupos a comparar, el test se puede aplicar con dos observaciones por grupo [32].

II.5. Detección de valores extremos

II.5.1. Valores extremos: definición y tipos

Cuando se realiza un experimento, es posible encontrar, a pesar de una buena planificación del mismo, resultados que puedan parecer anómalos con respecto al resto o a los esperados, es decir, valores extremos [33]. Según la norma ISO3534[34], los valores extremos se definen como observaciones de una muestra cuyo valor aparece bastante separado del resto, lo que sugiere que puede provenir de una población diferente a la estudiada o son el resultado de un error en la medida. Estos valores extremos se pueden dividir en dos grupos: valores aberrantes (outliers) cuando se detectan para un nivel de confianza del 99% y valores rezagados (stragglers) cuando son detectados entre el 95 y el 99% del nivel de confianza [35].

Cuando en una serie de resultados tenemos observaciones aberrantes, es necesario eliminarlas, para evitar la distorsión de las conclusiones. Eliminar los valores extremos implica el empleo de herramientas estadísticas, ya que aunque a priori pueden detectarse es necesario seguir unos criterios estadísticos fijados. Los resultados outliers, pueden aparecer en los valores promedio y en la varianza de los replicados siendo outliers centrales en el primer caso, y outliers dispersivos en el segundo. Existen varios

criterios para establecer los límites para los cuales un valor se considera extremo o no, pero en este caso se explicará la forma de detección de outliers centrales según la AOAC [36].

II.5.2. Detección de outliers centrales.

Cuando se sospecha que un valor se aleja del resto, hay que realizar un ensayo estadístico para comprobar si efectivamente es significativamente diferente del resto. Como con cualquier prueba de significación es necesario calcular un estadístico y comparar con los valores críticos tabulados para un nivel de confianza dado. Si los valores calculados son mayores que los tabulados, los valores sospechosos se considerarán outliers o stragglers, según el criterio establecido anteriormente.

Existen diversos ensayos para el contraste de outliers centrales, y cada uno depende del tipo de outlier que se trate: un solo valor sospechoso correspondiente al primero o último valor de los resultados ordenados, dos valores sospechosos y en extremos opuestos o dos valores extremos alejados del resto, pero vecinos entre sí. Para ello, la AOAC recomienda los ensayos de Grubbs con los que se pueden detectar outliers en estos tres casos [37]. Los valores tabulados se muestran en la tabla A2.8 del anexo 2.

- *Ensayo G1 para outliers simples*

Mediante este ensayo se puede comprobar si un valor sospechoso es anómalo, comparándolo con la media de los resultados. El cálculo del estadístico es el siguiente:

$$G1 = \frac{|\bar{x} - x_i|}{s} \quad (2.31)$$

siendo x_i el valor sospechoso, \bar{x} la media del conjunto global de valores y s la desviación estándar.

- *Ensayo G2 para outliers dobles separados diametralmente*

Cuando dos valores se encuentran distantes del resto pero en extremos opuestos, el estadístico G2 se calcula según:

$$G2 = \frac{x_n - x_l}{s} \quad (2.32)$$

siendo x_n el valor extremo superior de la serie, x_l el valor extremo inferior y s la desviación estándar del conjunto de resultados.

- *Ensayo G3 para outliers dobles vecinales*

El estadístico G3, que se calcula según la ecuación 2.33, sirve para comprobar si dos valores extremos vecinos se deben considerar outliers.

$$G3 = 1 - \left[\frac{(n-3)s_{n-2}^2}{(n-1)s^2} \right] \quad (2.33)$$

Donde n es el número total de valores, s^2 su varianza y s_{n-2}^2 la varianza de los valores excluyendo los extremos.

II.5.3. Detección de outliers dispersivos

Cuando se presenta una serie de resultados agrupados en replicados, y se sospecha que un nivel de replicación puede constituir un outlier dispersivo, es decir, que presenta una varianza superior a la del resto de series, se usa la prueba de Cochran[38, 39]. La hipótesis nula que se propone es que todos los grupos provienen de una población de igual varianza. El estadístico C de Cochran se calcula según la siguiente fórmula:

$$C = \frac{S_{max}^2}{\sum_{i=1}^l S_i^2} \quad (2.34)$$

siendo S_{max}^2 la varianza sospechosa, l el número de grupos y S_i^2 las varianzas de cada una de estas series. El valor obtenido se compara con el tabulado $C(\alpha, l, n)$ para un nivel de significación α y n replicados. El valor tabulado se obtiene como

$$C(\alpha, l, n) = \frac{1}{1 + (l-1)f} \quad (2.35)$$

siendo f el valor del estadístico F de Fisher para una probabilidad $\alpha' = 1 - \alpha/l$, para $(l-1)(n-1)$ y $(n-1)$ grados de libertad (tabla A2.9. del anexo 2).

Aunque este criterio está pensado para un número de replicados fijo para cada serie ($n = \text{constante}$), se puede modificar el procedimiento para aplicarlo de forma general. Simplemente se calcula un valor promedio del número de medidas y se compara con el valor tabulado $C(\alpha, l, \bar{n}, \dots)$, siendo:

$$\bar{n} = \frac{\sum_{i=1}^l n_i}{l} \quad (2.36)$$

II.6. Diseño experimental

La mayoría de los métodos estadísticos de diseño experimental se deben a trabajos desarrollados por R. A. Fisher y F. Yates durante estudios sobre actividades agrícolas [40]. Se entiende por diseño experimental, el proceso de plantear los pasos que se deben seguir en una investigación, y en qué orden para obtener la máxima información con el coste mínimo y la eficiencia máxima. Por ello, hoy en día el diseño experimental está ligado a multitud de campos disciplinares como por ejemplo el campo industrial o de análisis químico [41]. Montgomery [42] describe siete fases que definen un correcto diseño de experimento y se recogen en el esquema de la tabla II.2.

Si nos centramos en el campo de la química, cuando se realiza una investigación, existen factores (variables independientes) que se pueden controlar, como la temperatura o la ejecución por parte de un analista u otro, y se llaman factores controlados. También existen otros factores que no se pueden controlar denominados factores incontrolables. El modelo de la superficie de diseño puede escribirse como:

$$Y = \Phi(x_1, x_2, \dots, x_i) + e \quad (2.37)$$

donde x_i son los distintos factores que pueden afectar al experimento y e el error experimental.

El número de experimentos posibles, cuando se consideran todas las combinaciones de niveles de cada factor, de un diseño experimental completo, viene dado por la expresión:

$$N = n^f \quad (2.38)$$

donde n es el número de niveles y f el número de factores.

Tabla II.2. Fases del diseño experimental según Montgomery.

1. Reconocimiento y formulación del problema.	Preguntar por qué se va a realizar la investigación o qué conclusiones se pretenden sacar.
2. Selección de los factores y niveles.	Seleccionar los factores que se van a analizar, especificar si son cuantitativos o cualitativos y escoger los niveles que se van a estudiar. Los niveles pueden ser fijos o aleatorios. El número de niveles y factores van a influir en el tamaño del diseño.
3. Selección de variables de respuesta.	Elegir qué variables de respuesta o variables dependientes son las que proporcionan la información adecuada y definir cómo se van a medir.
4. Selección del diseño experimental	Elegir el tipo de diseño experimental que conduzca a resultados óptimos en función del número de niveles y factores.
5. Realización del experimento.	Recolectar los datos experimentales
6. Análisis de los datos.	Tratamientos estadísticos de los datos obtenidos
7. Conclusiones y recomendaciones	Decidir sobre las condiciones de los factores, teniendo en cuenta, también, criterios de practicidad.

Existen multitud de diseños experimentales en función del objetivo o estrategia de comparación y la forma de variar las variables independientes. Algunos de los diseños más utilizados se resumen en la figura II.4., y se detallan a continuación.

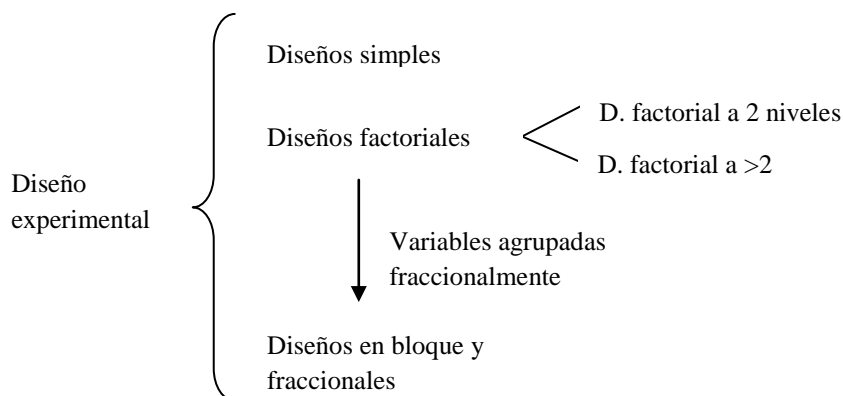


Figura II.4. Esquema de clasificación de tipos de diseños experimentales.

II.6.1. Diseño experimental simple

En un diseño simple, se aplica un valor conocido de una variable independiente a un grupo de muestras y otro valor de la misma variable independiente a un segundo conjunto de muestras y se estudia mediante la función de respuesta o variable dependiente el efecto de dicha variación. Para evaluar los resultados obtenidos es frecuente el uso de ANOVA o métodos de regresión lineal múltiple, MRL [43].

II.6.2. Diseño experimental factorial

En determinadas ocasiones es importante estudiar la influencia de varias variables independientes y para ello es necesario definir niveles dentro de las variables para poder estudiar las posibles interacciones. El caso más simple es el diseño experimental factorial a dos niveles.

II.6.2.1. Diseño factorial a dos niveles.

Un diseño experimental a dos niveles es de gran utilidad, ya que se evalúa la influencia de las variables independientes, pero con la ventaja de un número de experimentos no demasiado grande, 2^f , con cada factor a dos niveles. En primer lugar es necesario codificar los niveles y se suele hacer según: nivel alto = +1, y nivel bajo = -1. El hecho de codificar las variables introduce la ventaja de un sistema de coordenadas adimensional de los factores. Por tanto, en el espacio de factores los niveles más altos son +1, los más bajos -1 y las coordenadas del centro del diseño son cero y coinciden con el origen de coordenadas. Esta codificación es equivalente a reemplazar cada factor por su *z-score*.

$$z_i = \frac{x_i - x_i^0}{\Delta x_i} \quad (2.39)$$

$$x_i^0 = \frac{x_i^{\text{alto}} + x_i^{\text{bajo}}}{2} \quad (2.40)$$

$$\Delta x_i = \frac{x_i^{\text{alto}} - x_i^{\text{bajo}}}{2} \quad (2.41)$$

Otra de las ventajas que se obtienen al codificar los factores, es que se construye una matriz de diseño mediante vectores (columnas) ortogonales entre sí que simplifican los análisis de regresión. La matriz es de filas por columnas tiene la forma $2^f \times f$ de los niveles codificados.

Si se tienen tres factores ($f=3$) codificados a dos niveles ($n=2$) será necesario ejecutar $2^3=8$ veces el experimento para tener el diseño completo. La matriz de diseño que contiene 8×3 de niveles codificados se presenta en la tabla II.3. En dicha tabla se observan todas las combinaciones posibles para tres factores a dos niveles y un nivel correspondiente al cero (0).

Tabla II.3. Matriz de diseño para tres factores x_1, x_2, x_3 , con dos niveles y ocho experimentos.

Nº exp	Factor			Respuesta
	x_1	x_2	x_3	Y
1	+	+	+	Y_1
2	+	+	-	Y_2
3	+	-	-	Y_3
4	+	-	+	Y_4
5	-	-	+	Y_5
6	-	+	-	Y_6
7	-	-	-	Y_7
8	-	+	+	Y_8
0	0	0	0	Y_9

El modelo estadístico de un diseño factorial a dos niveles es lineal. Estos modelos representan superficies más o menos curvadas debido a los términos de interacción entre factores, pero sin llegar a máximos y mínimos y se llaman modelos de primer orden [44]. La función completa para tres factores según el modelo de regresión lineal múltiple es:

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \quad (2.42)$$

Los coeficientes b_1, b_2 y b_3 consideran los efectos principales de los factores, x_1, x_2 y x_3 , respectivamente. Por otro lado, b_{12}, b_{13}, b_{23} y b_{123} representan los términos de

las tres interacciones de segundo y tercer orden, respectivamente. El término b_0 representa la respuesta en el nivel cero de cada factor. Si los factores no tuvieran ningún efecto, las respuestas se distribuirían alrededor de b_0 con un error experimental debido al ruido. Este tipo de diseños a dos niveles, se denominan diseños saturados porque el número de ejecuciones requerido es igual al número de coeficientes a determinar. Sólo en el caso de realizar replicados o experimentos adicionales, se podrá estimar el error. En caso contrario no será posible porque no existen grados de libertad, ya que para realizar una regresión el número de grados de libertad necesarios es igual al número de ensayos realizados menos el número de coeficientes a determinar [43].

A partir de los valores de las respuestas, Y , se pueden estimar los coeficientes, pero a falta de grados de libertad, no se pueden estimar los errores asociados a los mismos. La interpretación de dichos valores deberá realizarse empleando el sentido común, eliminando los resultados próximos a cero. Si se consigue comprobar que algunos coeficientes no son significativamente diferentes de cero, entonces se pueden eliminar y ya se tienen grados de libertad para poder calcular errores asociados a coeficientes. En este caso, se realiza un ensayo de significación para los coeficientes de la regresión, del tipo t de Student.

$$t_k = \frac{|b_k|}{s_{bk}} \quad (2.43)$$

siendo b_k el coeficiente determinado y s_{bk} su desviación estándar asociada. Un coeficiente será significativo si el valor de t_k es mayor que el $t_{\text{tab}}(v, \alpha)$ para v grados de libertad y un nivel de significación α .

La única manera de poder determinar los coeficientes sin obviar ninguno de ellos, es realizar replicados de los experimentos, bien realizando duplicados de cada experimento o replicar en el centro del diseño, obteniendo 2×2^f y $2^f + r$ respuestas, respectivamente. Esto es equivalente a decir 2^f y r grados de libertad. En cualquier caso, es necesario aplicar una prueba t de Student para eliminar los coeficientes no significativos y recalcular el modelo.

II.6.2.2. Diseño factorial a más de dos niveles.

Cuando f es muy grande, el número de observaciones que se necesita para completar un diseño factorial, 2^f , es demasiado elevado. Por ello, surgen los diseños factoriales fraccionados del tipo 2^{f-p} , en los que se necesita agrupar las variables en bloques.

II.6.3. Diseño experimental de bloque y diseño fraccional

Uno de los requisitos primordiales para la metodología de superficie de respuesta es que los factores incontrolables deben producir como salida un ruido de fondo que varíe de forma aleatoria. Para poder asegurar esta condición es necesario que los experimentos se realicen bajo las mismas condiciones, lo cual es imposible si son demasiados. Por ello, en este diseño, el experimentador agrupa las unidades experimentales o experimentos en bloques. Los experimentos que pertenecen a cada bloque se suelen escoger según su homogeneidad (entre ellos), pero diferentes en término medio a otros bloques [45].

El diseño en bloques más simple es en el que cada factor se observa el mismo número de veces en cada bloque y se denomina diseño en bloques completo. La matriz de un diseño completo para cuatro factores x_1 , x_2 , x_3 y x_4 , con dos niveles y 8 experimentos, siendo $x_4 = x_1x_2x_3$ la variable de bloque, se presenta en la tabla II.4. Cuando $x_1x_2x_3 = 1$ se tiene el bloque I (experimentos 1, 3, 5 y 6) y cuando $x_1x_2x_3 = -1$ se tiene el bloque II (experimentos 2, 4, 7 y 8). Cada uno de estos bloques constituye una media fracción (2^{3-1}) del diseño global 2^3 .

Tabla II.4. Matriz de diseño completo para cuatro factores x_1 , x_2 , x_3 y x_4 , con dos niveles y 8 experimentos, siendo $x_4=x_1x_2x_3$ la variable de bloque.

Nº exp.	Factor				Respuesta
	x_1	x_2	x_3	$x_4=x_1x_2x_3$	Y
1	+	+	+	+	Y_1
2	+	+	-	-	Y_2
3	+	-	-	+	Y_3
4	+	-	+	-	Y_4
5	-	-	+	+	Y_5
6	-	+	-	+	Y_6
7	-	-	-	-	Y_7
8	-	+	+	-	Y_8

Uno de los fundamentos para el uso de fracciones es la hipótesis de que las interacciones de orden alto no son significativas. Así, se resuelve de manera sencilla las confusiones (cuando los efectos de doso más factores no se pueden separar). Esto se basa en la idea de que cuanto mayor es el orden de una interacción menor es su efecto y, por tanto, una fracción es buena si se confunden los efectos principales de los factores con interacciones de órdenes altos. Siguiendo el mismo criterio, una fracción es mala si los efectos principales de los factores son confundidos entre sí. La manera de cuantificar estas diferencias en el tipo de confusión que se dan en una fracción, es mediante un concepto importante que surge en estos casos y se denomina resolución del diseño. Se define la resolución como [46]:

$$R = I + 1 \quad (2.44)$$

donde I es el orden de interacción más bajo confundido con algún efecto principal, de forma que a mayor resolución mejor sea la fracción. En un diseño de resolución completa, se separan las interacciones y los efectos principales. Cuando la resolución es V, implica que se separan las interacciones binarias y efectos principales. En un diseño de resolución IV, se confunden entre sí las interacciones de segundo orden pero se separan de los efectos principales. Y en un diseño de resolución III, se confunden los efectos principales con las interacciones de segundo orden.

Existen diseños de bloque fraccionales de órdenes superiores, como un cuarto 2^{f-2} , un octavo, 2^{f-3} y en general, un k-avo, 2^{f-k} . Un diseño 2^{f-k} es equivalente a agrupar 2^k ejecuciones en 2^k bloques seleccionando k variables de bloque. El tipo de diseño completo o fraccional junto con el número de experimentos a realizar en este tipo de diseños 2^{f-k} se resumen en la tabla II.5.

Tabla II.5. Resolución y número de experimentos necesarios para diseños completos y fraccionales 2^{f-k} .

f	k	Diseño	Resolución	Nºexp.
3	0	2^3	Completa	8
3	1	2^{3-1}	III	4
4	0	2^4	Completa	16
4	1	2^{4-1}	IV	8
5	0	2^5	Completa	32
5	1	2^{5-1}	V	16
5	2	2^{5-2}	III	8
6	0	2^6	Completa	64
6	1	2^{6-1}	VI	32
6	2	2^{6-2}	IV	16
6	3	2^{6-3}	III	8
7	0	2^7	Completa	128
7	1	2^{7-1}	VII	64
7	2	2^{7-2}	IV	32
7	3	2^{7-3}	IV	16
7	4	2^{7-4}	III	8

II.6.4. Evaluación de los resultados de un diseño experimental: Función deseada

Existen diferentes métodos para evaluar un diseño experimental en términos de decisión sobre un conjunto de variables dependientes. Estos métodos se denominan MCDM, del inglés multi-criterion decisión-making.

La función deseada de Derringer es un método de toma de decisiones basado en varios criterios propuesto por Harrington [47] y posteriormente desarrollado por Derringer en el campo de la calidad en el desarrollo de productos, aunque es una aproximación a los MCDM muy empleada en el campo de la química y en particular a la optimización de métodos de análisis por HPLC [48]. Esta función es una medida global de la calidad y proporciona información para comparar y elegir los parámetros que optimicen el análisis.

El cálculo de la función deseada se basa en la transformación de las propiedades medidas en una escala adimensional, d , de manera que aunque se evalúen parámetros medidos en distintas unidades, se pueden combinar [49]. La escala de la función deseada va desde $d = 0$, para una respuesta no deseada a $d = 1$, para una respuesta totalmente deseada, y por encima de la cual una mejora de las variables dependientes no generan mejoras globales. Existen dos tipos de transformaciones posibles, dependiendo de los valores óptimos se encuentran en un los límites de una serie de valores o si se encuentran en un intervalo. En el primer caso, y cuando se pretende maximizar el resultado de una variable se realiza una transformación unilateral aplicándose la siguiente ecuación,

$$\left. \begin{array}{ll} d = 0 & \text{si } Y_i \leq Y^- \\ d = \left(\frac{Y_i - Y^-}{Y^+ - Y^-} \right) & \text{si } Y^- < Y_i < Y^+ \\ d = 1 & \text{si } Y_i > Y^+ \end{array} \right\} \quad (2.45)$$

donde Y^- es el valor mínimo aceptado según un criterio establecido e Y^+ es el resultado máximo por encima del cual no se produce ninguna mejora. Estos parámetros deben ser fijados por el usuario. Cuando se quiere minimizar el valor de una variable, nos encontramos con un caso de transformación unilateral inversa donde las ecuaciones que se aplican son las del conjunto 2.45, pero en lugar de buscar la forma de maximizar los valores, se pretende minimizarlos.

En el caso de que no interesen valores extremos sino valores de dentro de un intervalo, se aplica una transformación bilateral, dada por las siguientes ecuaciones [50],

$$\left. \begin{aligned} d &= \left(\frac{Y_i - Y^-}{c - Y^-} \right) \quad \text{si } Y^- \leq Y_i \leq c \\ d &= \left(\frac{Y_i - Y^+}{c - Y^+} \right) \quad \text{si } c < Y_i \leq Y^+ \\ d &= 0 \quad \text{si } Y_i < Y^- \text{ o } Y_i > Y^+ \end{aligned} \right\} \quad (2.46)$$

siendo c un valor deseado que se encuentra en el intervalo de $[Y^-, Y^+]$

El siguiente paso es obtener un valor global de D como la media geométrica de los distintos valores d obtenidos dependiendo de los criterios establecidos. Si el valor D toma valores en el intervalo $[0-1]$, es que todos los parámetros a estudiar han sido optimizados correctamente, mientras que si toma valores por debajo de $D = 0$, es que hay parámetros cuyos resultados están por debajo del límite deseado [51].

II.7. Reconocimiento de patrones

II.7.1. Introducción a los métodos de reconocimiento de patrones

El reconocimiento de patrones es una rama de la inteligencia artificial que aborda problemas de clasificación a partir de grandes conjuntos de datos [52]. Su campo de aplicación es bastante extenso, abordando problemas de clasificación de disciplinas como la psicología, la medicina, la química o la lingüística, entre otras. Los grandes avances informáticos contribuyen de forma directa al rápido avance de las técnicas de reconocimiento de patrones.

Al hablar de reconocimiento de patrones, el término patrón se refiere a un tipo de tema, suceso u objeto recurrente de un conjunto de objetos. Estos elementos se repiten de manera predecible y presentan características bien establecidas y definidas. A un conjunto de patrones caracterizados por una relación común se le denomina *clase*. En el caso de la química, los patrones son conjuntos de muestras descritas por una serie de variables o descriptores, que pueden ser parámetros físicos o químicos y que

caracterizan a dichas muestras. En reconocimiento de patrones se puede utilizar tanto métodos de análisis multivariante, como otros que no se basen en la estadística, por ejemplo, algoritmos neuronales.

Albano et al. [53], definieron cuatro niveles de reconocimiento de patrones. El primer nivel consiste en establecer fronteras entre patrones que permitan la clasificación de muestras dentro de esas clases de patrones previamente definidas. Se asume, por tanto, que todos los objetos pertenecen a alguna de las clases inicialmente definidas. Cuando, como en este caso, el patrón va asignado a una determinada clase, hablamos de “modelado duro” (*hard modelling*). El segundo nivel añade al primero la posibilidad de outliers. Es decir, si un objeto está dentro de la clase definida por una estructura matemática cerrada, se considera perteneciente a la clase. Los objetos fuera de estas fronteras se considera que no pertenecen a ninguna de las clases definidas. Se trata de un “modelado suave” (*soft modelling*). En el tercer y cuarto nivel propuesto por Albano et al. se estudia la capacidad para relacionar las variables medidas con otras variables o propiedades externas.

Dentro del campo del análisis y la composición de alimentos podemos considerar varios factores básicos desde el punto de vista de la diferenciación de clases. En primer lugar la aplicación de la ingeniería genética a los alimentos procedentes del mundo animal y vegetal. El segundo factor es el componente ambiental, considerando el tipo de suelo, clima y organismos simbióticos o parásitos. El tercer factor es el componente agronómico, que tiene en cuenta las prácticas de cultivo, riego, fertilizantes, abonos, piensos, medicamentos, pesticidas y recolección. El cuarto factor es el sistema de elaboración utilizado para obtener el producto final [54]. De esta forma, se suelen considerar los problemas de clasificación atendiendo, principalmente, al origen geográfico de los productos, a las variedades genéticas o a los distintos procedimientos de elaboración de los mismos.

II.7.1.1. Conceptos básicos

Espacio de patrones (Pattern Space)

Cualquier objeto i estudiado, en este caso hablamos de muestra, viene descrito por un conjunto c de variables y se representa por un vector fila:

$$x_{ij} \quad (j = 1 \text{ a } c) \quad (2.47)$$

cuyas componentes serán $x_{i1}, x_{i2}, \dots, x_{ic}$. Estos vectores que forman las filas de la matriz de datos \mathbf{X} en el espacio de las variables reciben el nombre vectores patrón (pattern vector). Si trabajamos en el espacio de los patrones u objetos, hablamos de vectores columna:

$$x_{ij} \quad (i = 1 \text{ a } r) \quad (2.48)$$

cuyas componentes son los valores de una misma variable en cada una de las muestras $x_{1j}, x_{2j}, \dots, x_{rj}$, y forman las columnas de la matriz de datos.

Estas formas de analizar los datos conducen a dos categorías de técnicas analíticas llamadas modo R y modo Q. Las técnicas en modo R tratan las relaciones entre las variables del experimento y examinan las dependencias entre las columnas de la matriz de datos (vectores columna), mientras que las técnicas en modo Q tratan las relaciones o agrupamientos entre los casos examinando las dependencias entre las filas de la matriz \mathbf{X} (vectores patrón).

Cuando se trabaja en el modo R, el punto de partida es calcular la matriz de covarianzas \mathbf{C} , que se obtiene según $\mathbf{C} = \mathbf{X}^T \mathbf{X}$. Es necesario el autoescalado antes del cálculo de la matriz de covarianzas con el fin de evitar el predominio de descriptores que presenten valores elevados sobre otros de valores más pequeños. Este autoescalado se realiza según:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (2.49)$$

donde:

$$\bar{x}_j = \frac{\sum_{i=1}^r x_{ij}}{r} \quad (2.50)$$

$$s_j = \sqrt{\frac{\sum_{i=1}^r (x_{ij} - \bar{x}_j)^2}{r-1}} \quad (2.51)$$

Así la matriz de covarianza para los datos autoescalados se transforma en la matriz de correlación \mathbf{R} , donde los elementos de la diagonal r_{ij} son siempre la unidad y el resto de elementos son los coeficientes de correlación $r_{ij} = r_{ji} \leq 1$.

Si se trabaja en modo Q, se comienza con una matriz de distancias en el espacio de los modelos ya que en este caso se estudian las relaciones entre los casos examinando las filas. Existen diversas formas de expresar la distancia entre dos vectores.

Distancias en el espacio patrón

La capacidad humana de encontrar patrones en el mundo externo es indiscutible ya que es nuestra forma de responder de una manera más sistemática a los estímulos externos [55]. En los años 30, R.A. Fisher [56] estableció los principios matemáticos de la estadística del reconocimiento de patrones, que es una de las maneras más adecuadas de afrontar el problema.

Se parte de que cualquier medida, objeto o patrón puede ser tomado como un punto en un espacio llamado espacio de patrones o espacio de modelos (Pattern Space). El espacio de modelos es, por tanto, el espacio de los datos de entrada. El conjunto de datos multivariante (N variables) puede representarse como un punto en un espacio de modelos multidimensional (N dimensional).

Cuando se trabaja en modo Q, examinando dependencias entre filas de la matriz \mathbf{X} hay que comenzar con una matriz de distancia en el espacio de modelos ya que la posterior clasificación se podrá efectuar en función de la menor distancia a una de las clases en el espacio de características. Si las clases están bien separadas entre sí, puede realizarse la clasificación en función de la distancia mínima del vector de

características a una de las clases. Se consideran dos vectores patrón \mathbf{X}_a y \mathbf{X}_b que pertenecen al espacio c -dimensional, de modo que,

$$\mathbf{x}_a = \{x_{a1}, x_{a2} \dots x_{ac}\} \quad (2.52)$$

$$\mathbf{x}_b = \{x_{b1}, x_{b2} \dots x_{bc}\} \quad (2.53)$$

Cuando se habla de distancia hay que referirse a la longitud del camino más corto entre dos entidades. Para un conjunto de elementos X , se define distancia como cualquier función binaria $d(a,b)$ de $X \times X$ en \mathbb{R} que verifique las siguientes condiciones:

a) No negatividad

$$d(a,b) > 0 \quad \forall a,b \in X \quad (2.54)$$

b) Simetría

$$d(a,b) = d(b,a) \quad \forall a,b \in X \quad (2.55)$$

c) Desigualdad triangular

$$d(a,b) \leq d(a,c) + d(c,b) \quad \forall a,b,c \in X \quad (2.56)$$

Existen varias formas de expresar las distancias cuantitativas pero las que más suelen emplearse son las siguientes:

- *Distancia euclídea*

Es una generalización del teorema de Pitágoras y a pesar de su sencillez de cálculo y de que verifica algunas propiedades interesantes tiene algunos inconvenientes. En primer lugar es que es una distancia sensible a las unidades de medida y, en segundo lugar, es que si las variables están correlacionadas la información que obtendríamos sería repetitiva. La distancia euclídea es recomendable cuando las variables son homogéneas y están medidas en

unidades similares y cuando se desconozca la matriz de varianzas. Se calcula según:

$$d_{ab} = \sqrt{\sum_{i=1}^c (x_{ai} - x_{bi})^2} \quad (2.57)$$

▪ *Distancia Mahalanobis*

En estadística, la distancia de Mahalanobis es una medida de distancia introducida por Mahalanobis en 1936. Su utilidad radica en que es una forma de determinar la similitud entre dos variables aleatorias multidimensionales y se diferencia de la distancia euclídea en que tiene en cuenta la correlación entre las variables aleatorias. Esta distancia solventa los dos inconvenientes que presenta el uso de la distancia euclídea, ya que es invariante ante los cambios de escala y no depende de las unidades de medida y elimina la posible correlación de las variables al introducir en el cálculo la matriz de covarianza, **C**. La distancia Mahalanobis viene dada por la fórmula de la ecuación 2.58.

$$d_{ab}^2 = (x_a - x_b)^T C^{-1} (x_a - x_b) \quad (2.58)$$

▪ *Distancia Minkowski*

Esta distancia presenta los mismos inconvenientes que la distancia euclídea. La distancia Minkowski se calcula según la ecuación 2.59, donde k es un número entero. Un caso particular de esta distancia es la distancia City Block o Manhattan, cuando k=1.

$$d_{ab}^2 = \sum_{i=1}^c (|x_{ai} - x_{bi}|^k)^{\frac{1}{k}} \quad (2.59)$$

▪ *Distancia City Block o Manhattan*

$$d_{ab}^2 = \sum_{i=1}^c |x_{ai} - x_{bi}| \quad (2.60)$$

- *Distancia de Canberra*

Es una modificación de la distancia City Block.

$$d_{ab}^2 = \sum_{i=1}^c \left| \frac{x_{ai} - x_{bi}}{|x_{ai}| + |x_{bi}|} \right| \quad (2.61)$$

- *Distancia de Karl Pearson*

Es una modificación de la distancia euclídea donde $\mathbf{S}_0=(s_1^2 \dots s_c^2)$ es la matriz diagonal que contiene las varianzas de $X_1 \dots X_c$. Esta expresión equivale a reescalar cada variable en unidades de desviación típica. El peso que se atribuye a la diferencia entre individuos es mayor cuanto menor es la dispersión en esa variable. Sigue suponiendo que las variables no están correlacionadas.

$$d_{ab}^2 = \sum_{i=1}^c \frac{(x_{ai} - x_{bi})^2}{s_i^2} = (x_a - x_b)' S_0^{-1} (x_a - x_b) \quad (2.62)$$

- *Distancia Hamming*

Es un caso particular de la distancia City Block para variables codificadas de forma binaria.

$$d_{ab}^2 = \sum_{i=1}^c \mathbf{XOR}(x_{ai}, x_{bi}) \quad (2.63)$$

II.7.1.2. Preprocesado de datos

El preprocesado de los datos sobre los que se va a trabajar consiste en realizar manipulaciones algebraicas sobre la matriz inicial, para mejorar la aplicación de las técnicas de reconocimiento de patrones. Existen, principalmente, dos métodos de preprocesado de datos como son: escalado y ponderación. En el escalado de datos se equiparan los valores de las variables, para evitar que existan distintos rangos de magnitud en las medidas, mientras que en el ponderado se pretende diferenciar entre los valores de descriptores que pertenecen a distintas clases, y su uso permite discernir entre las variables más importante a la hora de distinguir entre categorías.

En las técnicas de escalado se distinguen los siguientes procedimientos [57]:

- *Centrado*

Los nuevos valores vienen dados por la siguiente expresión

$$x = x_{ij} - \bar{x}_j \quad (2.64)$$

donde \bar{x}_j viene dado por la ecuación 2.50.

- *Normalización de columnas*

Se trata de dividir los valores de las variables entre la desviación estándar calculada según la ecuación 2.65.

$$x'_{ij} = \frac{x_{ij}}{s_j} \quad (2.65)$$

- *Autoescalado*

Las variables originales sufren una transformación tipo Student que viene dada por la ecuación 2.49.

- *Escalado del rango*

Transformación del intervalo de valores entre un máximo de 1 y un mínimo de 0.

$$x'_{ij} = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})} \quad (2.66)$$

- *Perfiles de fila*

consigue una frecuencia de aparición de cada variable.

$$m = \frac{1}{r} \sum_{i=1}^r x_i \quad (2.67)$$

En el caso de ponderación de las variables encontramos, principalmente, los siguientes procedimientos:

- *Ponderación de la varianza*[58]

se realiza aplicando la siguiente ecuación:

$$WV_j = \frac{r_1 (m_j^{(1)} - m_j)^2 + r_2 (m_j^{(2)} - m_j)^2}{S_j^{(1)2} + S_j^{(2)2}} \quad (2.68)$$

donde m es el valor promedio de todos los patrones considerados (ecuación 2.69) y $m^{(l)}$ el valor promedio de los patrones de una determinada clase L (ecuación 2.70). Cuanto mayor sea el valor de WV más discriminante será la variable j .

$$m = \frac{1}{r} \sum_{i=1}^r x_i \quad (2.69)$$

$$m^{(l)} = \frac{1}{r_l} \sum_{i=1}^{r_l} x_i^{(l)} \quad (2.70)$$

- *Pesos de Fisher* [59]:

$$FW_{j=} \frac{(m_j^{(1)} - m_j^{(2)})^2}{(s_j^{(1)})^2 - (s_j^{(2)})^2} \quad (2.71)$$

siendo $m^{(l)}$ el valor promedio de la clase L y $s^{(l)}$ su desviación estándar.

- *Ponderación de Coomans* [60], que es útil cuando se usan más de dos clases.

$$g_j = \frac{|m_j^{(1)} - m_j^{(2)}|}{s_j^{(1)} - s_j^{(2)}} \quad (2.72)$$

Siendo $m^{(l)}$ el valor promedio de la clase L y $s^{(l)}$ su desviación estándar.

- Método Λ de Wilks [61,62]

$$\Lambda = \frac{\det(\mathbf{S}_w)}{\det(\mathbf{S}_r)} \quad (2.73)$$

donde \mathbf{S}_r es la matriz de varianza total y \mathbf{S}_w es la matriz intraclase, que mide la dispersión dentro de cada clase. Los valores del parámetro Λ de Wilks van desde 1.0 (ningún poder discriminatorio) a 0,0 (poder discriminatorio máximo).

La matriz de varianza total es:

$$\mathbf{S}_r = \sum_{i=1}^r (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T \quad (2.74)$$

Para una variable, mientras que la covarianza entre las variables j y k es:

$$(\mathbf{S}_r)_{jk} = \sum_{i=1}^r (x_{ij} - m_j)(x_{ik} - m_k) \quad (2.75)$$

La matriz intraclase se define como:

$$\mathbf{S}_w = \sum_{l=1}^L \sum_{i=1}^{r_l} (\mathbf{x}_i^{(l)} - \mathbf{m}^{(l)})(\mathbf{x}_i^{(l)} - \mathbf{m}^{(l)})^T \quad (2.76)$$

Cuando es entre dos variables j y k se tiene:

$$(\mathbf{S}_w)_{jk} = \sum_{l=1}^L \sum_{i=1}^{r_l} (x_{ij} - m_j)(x_{ik} - m_k) \quad (2.77)$$

II.7.2. Análisis de componentes principales

El PCA es un método de visualización, que tiene como principal objetivo reducir el tamaño de la matriz de datos pero manteniendo la mayor parte de la información útil y descartando la información redundante. Cuando se tiene un conjunto de datos multivariantes en el que existen c variables medidas para el conjunto total de muestras y

estas variables están correlacionadas, la información que se tiene, como ya se ha mencionado anteriormente, es redundante. Se busca una nueva matriz, denominada matriz de correlación que se consigue transformando las c variables originales en otros c nuevos ejes llamados componentes principales, mediante un giro en el espacio de r dimensiones. Los nuevos ejes no están correlacionados entre sí, ya que son ortogonales y se van generando sucesivamente, de manera que el primero explica la mayor parte de la varianza total de los datos y los siguientes explican cantidades decrecientes de la varianza residual [63].

Matemáticamente expresado, se parte de una matriz de datos $\mathbf{X}_{r,c}$ sometida a un preprocesado. La matriz de covarianzas viene dada por $\mathbf{C}=\mathbf{X}^T\mathbf{X}$. El PCA transforma ortogonalmente esa matriz para tener una matriz de similaridad, $\mathbf{U}_{c,c}$ que actúe sobre la matriz de datos $\mathbf{X}_{r,c}$ para originar otra matriz $\mathbf{Y}_{r,c}$, llamada matriz de scores, en la que los datos están referidos a los nuevos ejes o PCs:

$$\mathbf{Y}_{r,c} = \mathbf{X}_{r,c}\mathbf{U}_{c,c} \quad (2.78)$$

La matriz $\mathbf{Y}_{r,c}$ debe cumplir:

$$\mathbf{Y}_{r,c}\mathbf{Y}_{r,c}^T = \mathbf{\Lambda}_{c,c} \quad (2.79)$$

donde $\mathbf{\Lambda}_{c,c}$ es la nueva matriz de covarianzas, que es una matriz diagonal. Por tanto se tiene que:

$$\mathbf{Y}^T\mathbf{Y} = (\mathbf{XU})^T (\mathbf{XU}) = \mathbf{U}^T\mathbf{CU} = \mathbf{\Lambda} \quad (2.80)$$

De donde se obtiene la expresión final $\mathbf{\Lambda} = \mathbf{U}^T\mathbf{CU}$ que se conoce como transformación de Karhunen-Loève (Karhunen-Loève transformation, KLT) [64]. \mathbf{U} es una matriz de transformación ortogonal, por lo que se cumple que $\mathbf{U}^T = \mathbf{U}^{-1}$. Si en la expresión de KLT se multiplica a ambos lados por \mathbf{U} se tiene que $\mathbf{U}\mathbf{\Lambda} = \mathbf{CU}$. Si se reescribe la ecuación matricial en forma vectorial se tiene que: $\mathbf{u}_j\lambda_j\mathbf{I} = \mathbf{CU}_j$. Esta es una ecuación de autovalores o autovectores, es decir, $(\mathbf{C}\lambda_j - \mathbf{I}) = 0$, por lo que debe cumplirse que $\det(\mathbf{C}-\lambda\mathbf{I})=0$, donde \mathbf{I} es la matriz unidad. Esto lleva a un polinomio de grado c :

$$\lambda^c + a_1\lambda^{c-1} + a_2\lambda^{c-2} + \dots + a_c = 0 \quad (2.81)$$

Como la matriz \mathbf{C} es simétrica las c soluciones son números reales y positivos. Los valores λ_j son elementos de la diagonal principal de la matriz $\mathbf{\Lambda}$ y son las varianzas de los datos referidas a los nuevos ejes. Se ordenan en sentido decreciente de los valores de λ_j , de forma que:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_c \quad (2.82)$$

siendo la suma de todos los λ_j igual a la suma de las varianzas de los datos originales. A partir de estos autovalores se calculan los vectores columna u_j y con ellos la matriz \mathbf{U} queda completa. Las coordenadas de los datos en los nuevo ejes, \mathbf{Y} , reciben el nombre de scores.

Las variables se pueden expresar linealmente en función de los PCs y viceversa. Así:

$$x_{ij} = a_{i1}Pc_{i1} + a_{i2}Pc_{i2} + a_{i3}Pc_{i3} + \dots + a_{ic}Pc_{ir} \quad (2.83)$$

los coeficientes a_{ik} se denominan pesos (loadings) e indican la contribución del componente principal PC_k a la variable original. Como los primeros PCs contienen la mayor parte de la información, es posible descartar aquellos que no proporcionen información relevante. La contribución a la varianza total de cada PC viene dada por la expresión:

$$\%var = \frac{\lambda_j}{\sum_{j=1}^c \lambda_j} \quad (2.84)$$

Algunos de los criterios empleados para seleccionar el número de componentes principales son:

- *Criterio de Kaiser* [64]

Es uno de los más sencillos y empleados. Admite que los datos están autoescalados y considera componentes explicativos aquellos cuyos autovalores correspondientes son mayores que la unidad ($\lambda > 1$).

- *Criterio de la proporción de la varianza explicada* [66]

Se trata de descartar componentes con autovalores muy pequeños. Normalmente, los componentes son eliminados hasta que el resto de ellos explican un porcentaje del total de la varianza de al menos un 70%.

- *Criterio de la validación cruzada* [67]

Se trata de un método basado en el algoritmo NIPALS (Nonlinear Iterative Partial Least Square). Básicamente este método se lleva a cabo dividiendo el conjunto de datos en un número determinado de grupos, normalmente de 5 a 9, y se van construyendo modelos con uno de los grupos no incluidos. Después de desarrollar todos los modelos se calcula la suma de cuadrados del error de predicción (Predicted Residual Sum of Squares, PRESS), como se indica en la ecuación 2.85, y se comparan mediante la ecuación 2.86, los errores PRESS de modelos que incluyen f componentes o $f+1$ componentes. Si la ecuación 2.86 toma valores >1 , no es necesario añadir más componentes, de lo contrario habrá que incluir uno a uno más hasta que el cociente sea mayor que uno.

$$PRESS_f = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.85)$$

$$\frac{PRESS_{f+1}}{PRESS_f} \quad (2.86)$$

donde \hat{y}_i es el valor predicho de la i -ésima respuesta observada e y_i el valor real de la i -ésima respuesta observada.

- *Selección escalonada*[68]

Los procesos de selección conocidos como *hacia atrás* o *hacia delante* y *por pasos*, son muy frecuentes ya que proporcionan buenos resultados cuando el número de variables es muy grande. En estos procedimientos las variables se introducen o eliminan del modelo siguiendo un criterio de entrada-salida. En el procedimiento *hacia delante* (forward) o de inclusión progresiva, se estudian las

correlaciones lineales simples de las variables con la respuesta y se incluye la que proporcione un mejor r^2 y que además cumplan con el criterio de entrada que es una prueba F:

$$F = \frac{SC'_{exp} - SC_{exp}}{\frac{SC'_{res}}{n - p}} \quad (2.87)$$

donde SC'_{exp} y SC_{exp} son las sumas de cuadrados de las varianzas incluyendo y excluyendo la variable estudiada, respectivamente y SC'_{res} es la suma de cuadrados de la varianza residual incluyendo la variable estudiada. En el procedimiento *hacia atrás* (backward) o de eliminación progresiva, se incluyen todas las variables y se van eliminando aquellas que están poco correlacionadas y que no cumplen el criterio F explicado en la ecuación 2.87. En el procedimiento *por pasos* (stepwise) se incluyen las variables como en el modelo de inclusión progresiva, sólo que una vez se han incluido, cabe la posibilidad de eliminar según el procedimiento de eliminación progresiva.

Cuando se han seleccionado los f componentes principales significativos mediante alguno de los criterios descritos, se pueden escribir las variables originales:

$$x_j = a_{j1}PC_1 + a_{j2}PC_2 + \dots + a_{jf}PC_f + e_j \quad (2.88)$$

siendo e_j el error de ajuste para la variable x_j .

Una vez realizado el PCA, se ha minimizado el número de variables sin eliminar información útil del conjunto de datos. PCA es, además, una manera efectiva de visualizar las relaciones existentes entre casos y descriptores, mediante el uso de BIPLOTS [69]. En este tipo de gráficos se representan a la vez, los scores de los objetos y las contribuciones de las variables con respecto a los primeros PCs. En la representación, las variables suelen ser vectores y los objetos puntos. La distancia euclídea entre dos puntos del BILOT es la distancia existente entre dos casos, por lo tanto con este tipo de representaciones se pueden estudiar posibles tendencias naturales

de los datos a agruparse en el espacio. En el caso de las variables, es posible decidir de manera gráfica, cuáles son las que mejor discriminan entre clases.

II.7.3. Reconocimiento de patrones no supervisados

El conjunto de técnicas que componen el grupo de reconocimiento de patrones no supervisados, se caracterizan porque ni las clases ni el número de clases se conocen a priori sino que se van definiendo conforme avanza la clasificación. Estas técnicas están basadas en agrupar los datos mediante algún criterio de proximidad, por lo tanto hay que definir:

- La medida de proximidad: indica como de similares o distintos pueden considerarse dos vectores de características.
- El criterio de agrupamiento: se define generalmente mediante alguna función de error que mide cómo de bueno son los agrupamientos.
- El algoritmo de agrupamiento: se encarga de obtener el óptimo global de la función error.
- El método de validación de los resultados: se encarga de verificar la validez de los resultados, normalmente mediante test estadísticos.

Un ejemplo de técnica de reconocimiento de patrones no supervisado es el análisis clúster cuyo objetivo es básicamente encontrar agrupamientos de muestras a partir de las variables medidas [70].

En este trabajo no se emplean técnicas de reconocimiento de patrones no supervisadas por lo que la atención se centrará en las técnicas de reconocimiento de patrones supervisadas.

II.7.4. Reconocimiento de patrones supervisados

Las técnicas de reconocimiento de patrones supervisadas suponen que se conoce a priori el número de clases así como la pertenencia a las mismas de cada uno de los miembros del conjunto de datos. La finalidad es diseñar y aplicar reglas de

clasificación para predecir las clases a las que pertenecen un conjunto de muestras desconocidas. Para llevar a cabo la clasificación el conjunto de muestras se divide en dos: un set de entrenamiento (training), sobre el que se diseña el modelo de clasificación y otro set de evaluación (test) [71]. La división suele ser de un 75% de los casos iniciales para el conjunto training y un 25% para el conjunto test [72]. La selección de estos porcentajes se realiza de manera aleatoria pero aplicándose a los objetos de cada clase, tomando estocásticamente el 75% de objetos en la clase I, la II, etc., para formar el conjunto de entrenamiento.

La habilidad predictiva del modelo se puede evaluar mediante validación cruzada de los siguientes tipos:

- *Validación cruzada aleatoria (Delete A Group Jacknife, DAGJK)* [73,74]: se trata de dividir aleatoriamente el conjunto de datos de entrenamiento y el conjunto de datos de prueba. Para cada subconjunto se calcula una función de clasificación que se ajusta a partir de los datos de training/test del subconjunto estudiado. El resultado final es el promedio de los valores obtenidos para las diferentes divisiones. Lo más común es utilizar la validación cruzada de 10 iteraciones.
- *Validación cruzada dejando uno fuera (leave-one-out)* [75]: En esta técnica se parte del conjunto conocido completo y a continuación se selecciona aleatoriamente uno de los objetos y *se deja fuera*, es decir, este único objeto va a constituir el conjunto de ensayo y el resto de casos formarán el conjunto de entrenamiento. Se desarrolla la regla de clasificación y se contabilizan los casos correctamente clasificados. Se repite el proceso de forma iterativa dejando un objeto distinto fuera. La eficacia predictiva se calcula contabilizando los éxitos y fallos en clasificar cada patrón.

Los fundamentos estadísticos comunes de los métodos reconocimiento de patrones supervisados se explican a continuación.

Considerando el caso particular de dos clases I y II, cada caso x_{ij} está representado por un vector fila, \mathbf{x} , cuyas componentes son los contenidos en cada

variable. Se denomina probabilidad a priori de la clase I, $P(I)$, a la probabilidad de que sin conocer el valor de las variables x , un caso pertenezca a la clase I. Si sólo se considera la existencia de dos clases se tiene que $P(I) + P(II) = 1$. La *probabilidad condicional* se define como $P(x/I)$, y es la probabilidad de que los valores de las variables de un caso que pertenece a la clase I sean los componentes de x . Según los axiomas de probabilidad propuestos por Kolmogorov en 1933, donde se establecen las condiciones mínimas que deben verificarse para que una función definida sobre un conjunto de sucesos determine consistentemente sus probabilidades [76], se tiene que:

$$P(x) = P(x/I)P(I) + P(x/II)P(II) \quad (2.89)$$

Por otro lado, se define probabilidad *a posteriori* como la probabilidad de que el caso pertenezca a la clase I conocido el valor de sus variables x . Según el teorema de Bayes:

$$P(I/x) = P(x/I)P(I)/P(x) \quad (2.90)$$

de manera análoga

$$P(II/x) = P(x/II)P(II)/P(x) \quad (2.91)$$

de acuerdo con esta aproximación estadística [77], se puede considerar que la región óptima, es decir la que minimiza la probabilidad de error, es la que asigna el caso a la clase para la cual la probabilidad *a posteriori* es mayor. Si $P(I/x) \geq P(II/x)$, el caso se asigna a la clase I.

II.7.4.1. Análisis Discriminante Lineal

Éste es uno de los métodos supervisados más aplicado, el cual estima la probabilidad a posteriori de que un objeto pertenezca a una determinada categoría, creando funciones de decisión que separan las clases a las que pertenecen los objetos a partir del conjunto de entrenamiento. El LDA consiste en encontrar las mencionadas funciones discriminantes lineales admitiendo que la separación entre categorías pueda realizarse mediante hiperplanos.

Esta técnica se basa en los trabajos desarrollados por Fisher a finales de los años treinta [78]. El procedimiento consiste en encontrar combinaciones lineales de las variables, que van a ser las denominadas funciones discriminantes, tal que sea máximo el cociente:

$$F = \frac{S_B}{S_w} \quad (2.92)$$

donde S_w es la varianza intramuestral, calculada según la ecuación 2.76, y S_B la varianza entre muestras, que se calcula como sigue:

$$S_B = \sum_{i=1}^L r_i \left(m^{(1)} - m \right) \left(m^{(1)} - m \right)^T \quad (2.93)$$

donde r_i y m quedan definidos en el apartado II.7.1.2.

Para realizar el cálculo de las funciones discriminantes, previamente, se seleccionan las variables con mayor poder discriminatorio utilizando para ello criterios de selección sucesiva (stepwise criteria) de forma que se incluyen sólo aquellas R variables que son verdaderamente relevantes para la discriminación [79]. A continuación se enumeran varios de estos criterios.

- Λ de Wilks [61], que ya fue introducido en la sección II.7.1.2. como:

$$\Lambda = \frac{\det(S_w)}{\det(S_T)} \quad (2.94)$$

Los valores de Λ van desde 1 (ningún poder discriminante) hasta 0 (máximo poder discriminatorio).

- Λ parcial de Wilks. Es la medida de Λ para la contribución de una variable determinada en el modelo:

$$\Lambda_{parcial} = \frac{\Lambda(\text{después de añadir la variable})}{\Lambda(\text{antes de añadir la variable})} \quad (2.95)$$

- Estadístico F [80], que se obtienen a partir del valor de Lambda parcial según:

$$F = \frac{N - k - R}{k - 1} \cdot \frac{1 - \Lambda_{parcial}}{\Lambda_{parcial}} \quad (2.96)$$

donde N es el número total de objetos, k es el número de clases y R el número de variables consideradas.

Si se van seleccionando las variables más discriminantes por un método de inclusión progresiva, al comienzo, en el paso 0, no hay variables en el modelo y, por definición, $\Lambda = 1$. Después de añadir la primera variable, se cumple que $\Lambda_{parcial} = \Lambda$.

De todas las variables, se selecciona como primera aquella cuyo valor de F tiene un menor nivel de significación (p-level). Se continúa añadiendo variables hasta que F presente un p-level superior a un valor especificado de antemano (por ejemplo 0,05). Existe un método análogo pero esta vez se incluyen de todas las variables en el modelo, eliminando una en cada paso, atendiendo a su menor poder discriminante y se denomina método de eliminación progresiva (backward selection).

Una vez seleccionadas las R variables más significativas, se procede al cálculo de las funciones discriminantes, también denominadas variables canónicas [81], que son combinación lineal de las R variables discriminantes, de modo que se observe la mayor separación posible entre clases y se reduzca el número de dimensiones sin que haya pérdidas en la diferenciación entre clases. Las funciones discriminantes constituyen unos nuevos ejes donde las diferencias entre categorías se observan más claramente. La expresión analítica de estas funciones es la siguiente:

$$f(x_1, x_2, \dots, x_R) = v_1 x_1 + v_2 x_2 + \dots + v_r x_R = v^T x \quad (2.97)$$

En definitiva, las funciones discriminantes son rotaciones en el espacio de las variables reducidas que generan combinaciones lineales que deben caracterizar bien una clase.

El objetivo de la técnica es hacer que los promedios entre cada categoría sean lo más diferentes entre sí y, por otra parte, que cada clase tenga la menor desviación

interna posible. Esto se consigue calculando los vectores \mathbf{v}^T que hacen máximo el cociente:

$$F = \frac{\mathbf{v}^T \mathbf{S}_B \mathbf{v}}{\mathbf{v}^T \mathbf{S}_w \mathbf{v}} \quad (2.98)$$

El vector \mathbf{v} , para el cual se maximiza el valor de F, será aquel para el cual la derivada primera sea nula:

$$\frac{\partial F}{\partial \mathbf{v}} = \frac{\partial(\mathbf{v}^T \mathbf{S}_B \mathbf{v} / \mathbf{v}^T \mathbf{S}_w \mathbf{v})}{\partial \mathbf{v}} = 0 \quad (2.99)$$

$$\frac{\partial F}{\partial \mathbf{v}} = \frac{\mathbf{S}_B \mathbf{v} (\mathbf{v}^T \mathbf{S}_w \mathbf{v}) - (\mathbf{v}^T \mathbf{S}_B \mathbf{v}) \mathbf{S}_w \mathbf{v}}{(\mathbf{v}^T \mathbf{S}_w \mathbf{v})^2} = 0 \quad (2.100)$$

de aquí:

$$\mathbf{S}_B \mathbf{v} (\mathbf{v}^T \mathbf{S}_w \mathbf{v}) - (\mathbf{v}^T \mathbf{S}_B \mathbf{v}) \mathbf{S}_w \mathbf{v} = 0 \quad (2.101)$$

$$\mathbf{S}_B \mathbf{v} - F \mathbf{S}_w \mathbf{v} = 0 \quad (2.102)$$

finalmente:

$$\mathbf{S}_w^{-1} \mathbf{S}_B \mathbf{v} = F \mathbf{v} \quad (2.103)$$

Esta es la típica ecuación de autovalores y autovectores. Así pues, habrá un número determinado de autovalores distintos de cero que corresponderán al número de funciones discriminantes para separar las categorías. Ese número, normalmente, es el mínimo entre el número de clases menos uno y el número de variables, es decir: $\min(k-1, R)$.

Por otra parte, se denomina *discriminant score* de un objeto X_i que pertenece a la clase k a la expresión:

$$DS_i^{(k)} = \mathbf{v}^T \mathbf{X}_i^{(k)} \quad (2.104)$$

De esta forma, el objeto X_i se puede clasificar según su posición con respecto al nuevo eje f , comparándolo con los centroides de las clases existentes. Suponiendo dos clases 1 y 2, cuyos centroides son, respectivamente, $\overline{DS}^{(1)}$ y $\overline{DS}^{(2)}$, el objeto X_i pertenecerá a la clase 1 si se cumple:

$$\left| DS_i - \overline{DS}^{(1)} \right| < \left| DS_i - \overline{DS}^{(2)} \right| \quad (2.105)$$

En caso contrario, pertenecerá a la clase 2.

Esta es una manera de clasificar objetos pero, generalmente en el análisis discriminante lineal, se suelen calcular reglas de clasificación basadas en la estimación de las probabilidades *a posteriori* de la pertenencia de un objeto a una clase empleando estimación Bayesiana.

De acuerdo con las definiciones de probabilidad *a priori*, probabilidad condicional y probabilidad *a posteriori*, expresadas en el apartado II.3.4. y según el teorema de Bayes, se tiene que la probabilidad *a posteriori*, para un objeto X_i de pertenecer a una clase c_k es:

$$P(c_k / x_i) = \frac{P\left(\frac{x_i}{c_k}\right)P(c_k)}{\sum_{j=1}^k P\left(\frac{x_i}{c_j}\right)P(c_j)} = \frac{P\left(\frac{x_i}{c_k}\right)P(c_k)}{P(x_i)} \quad (2.106)$$

Normalmente, si se desea establecer la regla de clasificación por el método bayesiano, se suelen hacer tres suposiciones:

- 1) Las características X_{ij} son estadísticamente independientes, lo cual significa que la matriz de covarianzas es diagonal con todos los elementos de la diagonal iguales.
- 2) Las distribuciones $P(X_{ij} | c_k)$ siguen la ley de Gauss.

3) Estas distribuciones tienen la misma varianza y solo se diferencian en sus valores promedio $\mu_j(I)$ y $\mu_j(II)$.

Así:

$$P\left(\frac{x_i}{c_1}\right) = \frac{1}{(2\pi)^{\frac{R}{2}} \sigma_1 \sigma_2 \dots \sigma_R} e^{-\frac{1}{2} \left[\left(\frac{x_{i1} - \mu^{(1)}}{\sigma_1^{(1)}} \right)^2 + \dots + \left(\frac{x_{iR} - \mu_R^{(1)}}{\sigma_R^{(1)}} \right)^2 \right]} \quad (2.107)$$

La expresión anterior se puede reescribir como:

$$P(x_i / c_1) = \frac{1}{(2\pi)^{R/2} (\det \Sigma)^{1/2}} e^{-\frac{1}{2} \left[(\mathbf{x}_i - \boldsymbol{\mu}^{(1)})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}^{(1)}) \right]} \quad (2.108)$$

donde:

$$\Sigma^{-1} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \sigma_R^2 \end{bmatrix} \quad (2.109)$$

En general:

$$P(x_i / c_k) = \frac{1}{(2\pi)^{R/2} (\det \Sigma)^{1/2}} e^{-\frac{1}{2} \left[(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}^{(k)}) \right]} \quad (2.110)$$

Tomando como función discriminante:

$$g(x_i)^{(k)} = \ln P(c_k) + \ln P(x_i / c_k) \quad (2.111)$$

Y sustituyendo $P(x_i / c_k)$ por su expresión, queda:

$$g(x_i)^{(k)} = \ln P(c_k) - \frac{1}{2} (R \ln 2\pi + \ln(\det \Sigma)) - \frac{1}{2} (\mathbf{x}_i^T - \boldsymbol{\mu}^{(k)})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}^{(k)}) \quad (2.112)$$

Los dos primeros términos de la expresión anterior son independientes de la clase k , ya que se admite que las probabilidades *a priori* son iguales $P(c_j) = P(k)$ y Σ es común. Por tanto:

$$g(x_i)^{(k)} = -\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)}) = -\frac{1}{2}D_i^{(k)^2} \quad (2.113)$$

donde $D_i^{(k)}$ es la distancia de Mahalanobis del patrón x_i al centroide de la clase k .

Operando, se tiene:

$$g(x_i)^{(k)} = -\frac{1}{2}\mathbf{x}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{x}_i + \mathbf{x}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^{(k)} - \frac{1}{2}\boldsymbol{\mu}^{(k)T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^{(k)} \quad (2.114)$$

El primer término de la expresión carece de poder discriminante, por lo que ésta se puede escribir como:

$$g(x_i)^{(k)} = \mathbf{x}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^{(k)} - \frac{1}{2}\boldsymbol{\mu}^{(k)T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^{(k)} \quad (2.115)$$

A esta expresión también se le denomina *classification score*, así:

$$CS_i^{(k)} = \mathbf{x}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^{(k)} - \frac{1}{2}\boldsymbol{\mu}^{(k)T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^{(k)} \quad (2.116)$$

En la práctica, se emplea S_w , es decir, la varianza intraclase, como una mejor estimación de Σ . Y, además, si llamamos $\mathbf{m}^{(k)}$ a la expresión:

$$\mathbf{m}^{(k)} = \frac{1}{r_k} \sum_{i=1}^{r_k} \mathbf{x}_i^{(k)} \quad (2.117)$$

El *classification score* queda:

$$CS_i^{(k)} = \mathbf{x}_i^T S_w^{-1} \mathbf{m}^{(k)} - \frac{1}{2}\mathbf{m}^{(k)T} S_w^{-1} \mathbf{m}^{(k)} \quad (2.118)$$

Esta es la denominada *función de clasificación de Fisher*, por la que cada objeto x_i se asigna a la clase k para la cual la función de clasificación $SC_i^{(k)}$ presente el mayor valor.

II.7.4.2. Máquinas de vectores soporte

Las máquinas de vectores soporte o SVM son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik [82]. Son métodos empleados en la resolución de problemas de clasificación y regresión. Cuando se tiene un conjunto de muestras de entrenamiento, se pueden etiquetar las clases y entrenar las SVMs para construir un modelo que prediga las clases de las muestras de un conjunto test [83]. Originariamente las SVMs fueron creadas para resolver problemas de clasificación binaria, y hoy en día sirven, también, para resolver problemas de regresión o multclasificación. Las SVM pertenecen a la familia de clasificadores lineales ya que se basan en separadores lineales o hiperplanos en espacios de alta dimensionalidad, pero también es posible la clasificación no lineal [84].

En concreto, Cortes y Vapnik [82] encontraron que si existía una frontera entre las regiones en las que se distribuyen las muestras de las distintas clases y la separación entre clases pudiera hacerse linealmente, se podría obtener un hiperplano óptimo que maximizara las distancias entre clases. SVM utiliza muestras de los márgenes de las clases (vectores soporte) para obtener un hiperplano óptimo que separe ambas clases. Dicho hiperplano equidista de los márgenes definidos por los vectores soporte. El número de vectores soporte debe ser pequeño, generalmente entre 2 y 4. En la figura II.5. se observa como los vectores soporte son los que deciden la posición del hiperplano y los puntos que no se encuentran en la frontera no tienen influencia en la posición del mismo. En el caso concreto de esta figura, se observa cómo se establece un hiperplano óptimo a partir de tres vectores soporte.

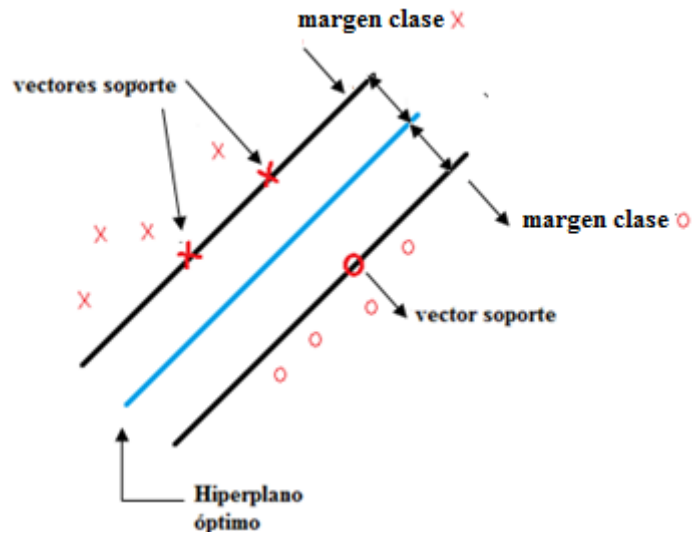


Figura II.5. Hiperplano óptimo a partir de tres vectores soporte.

Otra característica de las SVM es que, mientras la mayoría de los métodos de aprendizaje se centran en minimizar los errores cometidos por el modelo generado a partir de los casos de entrenamiento (error empírico), el sesgo inductivo asociado a las SVMs radica en la minimización del denominado riesgo estructural.

Las SVM permiten resolver problemas lineales basándose en separadores lineales o hiperplanos, pero en la realidad, los problemas no suelen tener solución lineal por lo que hay que recurrir a funciones matemáticas que permitan transformar el problema en un espacio de mayor dimensión, denominado espacio de características, donde sí es posible la separación lineal [84]. Esta transformación se realiza mediante funciones kernel.

Un modelo SVM pretende, por tanto, buscar una función adecuada que permita llevar a cabo una diferenciación de clases dentro de un conjunto de datos. La descripción de los algoritmos empleados en SVM se explica en tres partes. En primer lugar la definición básica para clases linealmente separables, en segundo lugar y con funciones kernel, la ampliación a casos no linealmente separables, y en tercer lugar la solución general con el término de error de penalización, que controla la complejidad del modelo [85].

En el caso de dos conjuntos linealmente separables, las SVM forman líneas o hiperplanos que separan los datos de entrada en dos grupos, cada uno con una etiqueta

propia $\{-1 \text{ y } 1\}$, y son estos datos los que definen la función de separación que se puede expresar de la siguiente manera:

$$g(x_i) = \text{sgn}(\mathbf{w}\mathbf{x}_i + b) = \text{sgn}(b + \sum_{j=1}^J w_j x_{ij}) \quad (2.119)$$

donde \mathbf{x} es el conjunto de datos de entrenamiento, \mathbf{w} y b son un vector peso y una constante (sesgo), respectivamente, determinados del conjunto de entrenamiento.

El signo de g determina a qué clase se asigna una muestra, por ejemplo A (>0) y B (<0). Existen muchos planos que satisfacen la condición $\langle \mathbf{w} \cdot \mathbf{x} \rangle + b = 0$, que es la que divide el espacio de datos en dos regiones de signo contrario. Si las dos clases se pueden separar, es posible definir un margen entre las dos clases de manera que:

$$\mathbf{w}\mathbf{x} + b \geq 1, \quad c = +1 \quad (2.120)$$

$$\mathbf{w}\mathbf{x} + b \leq -1, \quad c = -1 \quad (2.121)$$

Los parámetros \mathbf{w} y b se pueden escalar de manera adecuada para que siempre se tenga un margen con una distancia ± 1 de la línea central. La frontera o hiperplano, debe ser equidistante a los dos extremos de cada clase. Existen infinitud de hiperplanos o líneas que cumplen las condiciones para poder separar dos clases, por lo que hay que definir una regla de decisión para determinar cuál es el hiperplano óptimo. En la construcción de SVM se considera hiperplano óptimo, definido por \mathbf{w} y b , aquel que maximiza la distancia entre el hiperplano óptimo y el valor de entrada más cercano (maximización del margen) con la intención de forzar la generalización de la máquina que se esté construyendo.

La optimización se puede expresar de manera algebraica con la siguiente ecuación:

$$\varphi(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}(\mathbf{w}\mathbf{w}') - \sum_{i \in SV} \alpha_i (c_i [\mathbf{w}\mathbf{w}'_i + b] - 1) \quad (2.122)$$

donde el parámetro $\boldsymbol{\alpha}$ se denomina multiplicador de Larrange y se emplea para optimizar una función sujeta a una o más restricciones. El problema de optimización

del margen representa un problema de optimización cuadrático con restricciones lineales que se puede resolver mediante programación cuadrática.

El problema de dos clases linealmente separables no suele corresponderse a una situación real, ya que en los ejemplos reales suele haber ruido en las muestras y no son perfectamente separables por un hiperplano. Para ello se recurre a lo que se denomina margen máximo con holgura, permitiendo cierto grado de error en el hiperplano de separación de margen máximo. Desde el punto de vista de la formulación para definir el hiperplano óptimo, se incluyen unas variables reales denominadas variables de holgura, ξ_i , que permiten cuantificar el número de ejemplos no separables que se está dispuesto a admitir, es decir:

$$c_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, n. \quad (2.123)$$

De acuerdo a esta expresión, variables de holgura igual a cero corresponden a casos separables, mayores que cero corresponden a ejemplos no separables y mayores que uno corresponden a ejemplos no separables y mal clasificados. Como las restricciones se han relajado con la intención de poder separar ejemplos que no son perfectamente linealmente separables, en la expresión para conseguir el hiperplano óptimo, ya no basta con maximizar el margen, sino que hay que incluir los errores de clasificación que comete el hiperplano,

$$\varphi(\mathbf{w}, b, \xi) = \frac{1}{2} |w|^2 + C \sum_{i=1}^n \xi_i^p \quad (2.124)$$

donde C es una constante elegida por el usuario, denominada error de penalización [85], que permite regular el compromiso entre el grado de sobreajuste del clasificador y la proporción del número de ejemplos no separables. El nuevo problema de optimización consistirá en encontrar el hiperplano definido por \mathbf{w} y b , que minimicen la expresión 2.124 y sujeto a las restricciones de la ecuación 2.123. A los hiperplanos así definidos se les denomina de margen blando, en contraposición a los hiperplanos de margen duro que se definían para ejemplos perfectamente linealmente separables.

En la figura II.6 se representan los vectores soporte apoyados en los márgenes de cada clase, muestras dentro de los márgenes ($\xi=0$), muestras entre los márgenes ($0<\xi<1$) y muestras mal clasificadas $\xi>1$.

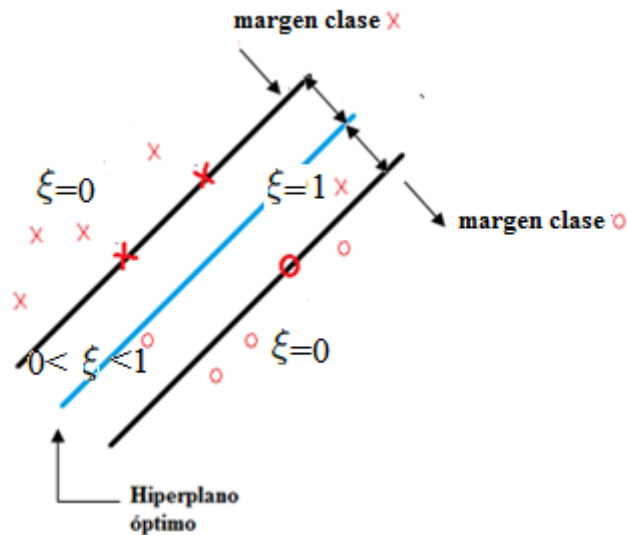


Figura II.6. Hiperplano óptimo a partir de tres vectores soporte, muestras clasificada correctamente y muestras mal clasificadas.

Hasta ahora se ha planteado la resolución de problemas de clasificación lineal o parcialmente lineal mediante SVM. No obstante, la mayoría de los problemas reales incluyen casos en los que la separación lineal no es posible de ninguna manera. Las SVM son más útiles en estos casos, ya que las separaciones lineales se pueden realizar mediante otras técnicas menos complejas, pero los problemas no linealmente separables no se pueden resolver con cualquier herramienta. En estos casos, la manera en que las SVM resuelven la situación es definir un espacio de dimensiones mayores al espacio de las variables, que se denomina espacio de características donde sí es posible la separación lineal. En la figura II.7. se representa esta transformación.

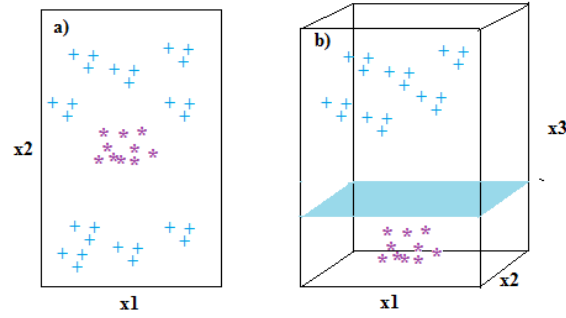


Figura II.7. Transformación del espacio de entrada a un espacio de características de mayor dimensión.

El principal inconveniente que encontramos es la dificultad de encontrar una función de transformación adecuada, además de la dificultad de transformar vectores \mathbf{x} (espacio de variables) en vectores \mathbf{f} (espacio de características). La solución a estos problemas son las funciones kernel.

El uso de funciones kernel permite describir de manera más sencilla el problema en un espacio de características de mayor dimensión y explicar un problema no lineal mediante métodos lineales. Las funciones kernel se aplican sobre vectores de X y su resultado es un producto escalar sobre un espacio de características F , es decir:

$$K(x_i, x_l) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_l) \rangle \quad K(x_i, x_l) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_l) \rangle \quad (2.125)$$

Las funciones kernel, por tanto, permiten calcular productos escalares en el espacio F de características aplicando la función kernel sobre el espacio X de las variables.

La optimización se escribe ahora de la siguiente manera,

$$\varphi(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{i \in VS} \sum_{l \in VS} \alpha_i c_i K(x_i, x_l) c_l \alpha_l - \sum_{i \in VS} \alpha_i \quad (2.126)$$

De nuevo, la tarea de optimización implica un problema de programación convexa cuadrática, que se resuelva de manera sencilla en términos de $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_l)$, ya que es posible la transformación sin tener que crear el espacio de características y trabajando solo en el espacio original en el que $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_l)$, es definido. En otras palabras, en un algoritmo que pueda expresarse en función de productos escalares sobre su espacio de entrada, se sustituye el producto escalar original por la función kernel. Se consigue que

el algoritmo original pase a aplicarse al espacio de características. Este atributo se conoce como el truco de kernel y es lo que hace que las SVM sean de gran utilidad en problemas de clasificación y regresión.

La función de decisión para la clasificación se formula de la siguiente forma:

$$g(\mathbf{x}) = \text{sgn} \left(\sum \alpha_i c_i \mathbf{K}(\mathbf{s}_i, \mathbf{x}) + b \right) \quad (2.127)$$

donde \mathbf{s}_i es una matriz que contiene los vectores soporte. El plano sigue dependiendo de los vectores soporte. Luego se vuelve el espacio de variables consiguiendo la separación mediante fronteras no lineales. En la figura II.8. se observa cómo se crea una frontera para un caso de muestras no separables. En un espacio de dimensiones mayores se genera un plano a partir de tres vectores soporte y cuando se vuelve al espacio inicial es posible la separación de las clases.

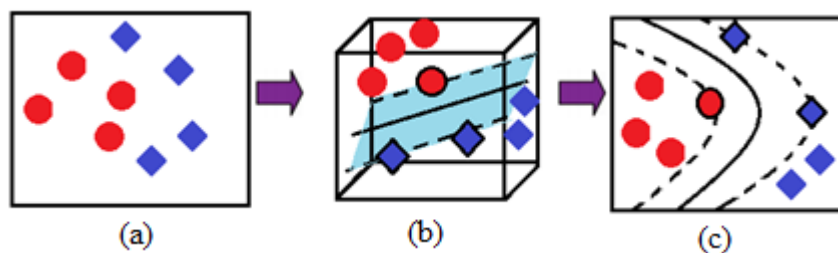


Figura II.8. (a) Clases no separables. (b) Proyección en un espacio de dimensiones superiores. (c) proyección en dos las dimensiones iniciales.

Las funciones kernel más comunes son las siguientes:

- Función de base radial (RBF) que se define como

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (2.128) \quad \text{o} \quad \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (2.129)$$

donde $\gamma=1/2\sigma$.

- Función polinómica (PF), $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = (a\mathbf{x}_j^T \mathbf{x}_i + b)^c$ (2.130).
- Función sigmoideal (SF), $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \tanh(a\mathbf{x}_j^T \mathbf{x}_i + b)$ (2.131).
- Función identidad, $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$ (2.132).

Las SVM presentan ciertas ventajas con respecto a otras técnicas de clasificación y regresión. Las SVM aprenden de los márgenes por lo que es posible trabajar con conjuntos de muestras pequeños, además transforman de espacios sin posibilidad de separación lineal en un espacio de dimensiones superior, donde la separación lineal es posible, luego vuelve al espacio de variables original generando fronteras no lineales. Es una técnica capaz de tratar problemas de alta dimensionalidad de manera satisfactoria [86].

Sin embargo, existe un inconveniente en el uso de las SVM y es el sobreajuste, que ocurre cuando se han aprendido muy bien los datos de entrenamiento pero no se pueden clasificar bien ejemplos nunca antes vistos. Para ello en la ecuación 2.124 se incluye un parámetro denominado error de penalización, basado en el principio de minimización del riesgo estructural, para que el usuario tenga en cuenta y optimice de manera que se evite en la medida de lo posible el sobreajuste.

II.7.4.3. Modelado suave independiente por analogía de clases.

El modelado suave independiente por analogía de clases (SIMCA) es una técnica supervisada de reconocimiento de patrones es un procedimiento denominado de modelización blanda, ya que puede que algún caso no se clasifique dentro de alguna de las categorías existentes, Así, el método SIMCA [53,87-90] asocia a cada clase conocida un modelo matemático individual, de tal forma que cada patrón desconocido se compara con cada uno de los modelos establecidos para comprobar si se ajusta a alguno de ellos, perteneciendo entonces a dicha categoría. En caso de pertenecer a otro tipo de distribución constituiría un outlier con respecto a las clases conocidas en el conjunto de entrenamiento.

SIMCA divide la matriz de datos original en submatrices correspondientes a cada una de las k categorías y realiza un PCA por separado para cada una de las clases del conjunto de entrenamiento [91]:

$$x^{(k)} = Y^{(k)}U^{T(k)} = y_1^{(k)}u_1(k) + y_2^{(k)}u_2(k) + \dots \quad (2.133)$$

Mediante el método de la validación cruzada y empleando el algoritmo NIPALS, explicado en el apartado II.7.2., se seleccionan f PCs explicativos de la mayor parte de la varianza de los datos para cada clase; así, se tiene:

$$x^{(k)} = \sum_{i=1}^{f^{(k)}} y_i^{(k)} u_i^{T(k)} + \mathbf{E} \quad (2.134)$$

donde \mathbf{E} es la matriz error que contabiliza los residuales de cada caso, es decir, la diferencia entre los datos originales y la estimación del modelo.

La varianza residual para la clase K con r_k objetos es:

$$s_k^2 = \frac{ss}{(r_k - f^{(k)} - 1)(v - f^{(k)})} \quad (2.135)$$

donde v es el número de variables originales y ss viene dada por la expresión:

$$ss = \sum_{i=1}^{r_k} \sum_{j=1}^f e_{ij}^2 \quad (2.136)$$

Para un elemento cualquiera, se tiene:

$$x_{kl}^{(k)} = \sum_{a=1}^{f^{(k)}} y_{ka}^{(k)} u_{al}^{(k)} + \varepsilon_{kl} \quad (2.137)$$

Con esta ecuación se calcula el valor estimado para x_{kl} según el modelo PCA realizado para la clase k y se evalúa el error de ajuste. Esto se repite para cada uno de los elementos en cada una de las clases.

Una vez evaluados los errores, la varianza para el elemento q será:

$$s_q^2 = \frac{\sum_{j=1}^v \varepsilon_{qj}^2}{v - f^{(k)}} \quad (2.138)$$

Si s_q^2 es de la misma magnitud que s_k^2 , el objeto q se considera miembro de la clase k. El criterio F de Fisher proporciona una medida cuantitativa para dicha clasificación:

$$F = \frac{s_q^2}{s_k^2} \quad (2.139)$$

Al emplear un criterio F para establecer la pertenencia de un objeto a la clase considerada, el método SIMCA es evidentemente paramétrico.

II.8. Regresión múltiple y multivariante

En muchas ocasiones nos encontramos con que es necesario predecir una variable respuesta a partir de otra variable relacionada y que se usa como predictora. Este problema es fácilmente abordable desde el punto de vista de la regresión lineal simple (RLS). Sin embargo, en otras ocasiones, es necesario tener en cuenta varias variables predictoras para generar una respuesta, lo cual se denomina regresión múltiple. En el caso de necesitar varias variables predictoras para varias variables respuesta, estaríamos hablando de regresión multivariante [92]. En este trabajo se utiliza la regresión múltiple en lugar de multivariante, y es por ello que se desarrollarán los conceptos relacionados con la primera.

II.8.1. Modelos de regresión múltiple

Cuando se pretende predecir una variable, y, a partir de variables independientes, $x_1, x_2 \dots x_i$ se emplea el siguiente modelo de regresión lineal múltiple, que es el más habitual en regresión lineal multivariante [93]:

$$\hat{y} = b_o + b_1x_1 + b_2x_2 + \dots b_jx_j \quad (2.140)$$

donde b_o es el valor de \hat{y} cuando $x_1, x_2 \dots x_i$ toman el valor cero, b_j indica la variación que se produce en \hat{y} por efecto de un aumento de x_j para todo x_i constante.

El siguiente problema que encontramos es que los coeficientes $b_1, b_2 \dots b_i$ no son comparables, ya que las variables $x_1, x_2 \dots x_i$ no están medidas en las mismas unidades y no tienen los mismos errores. Por ello, se describen los términos β_i o coeficientes de correlación, para comparar el aporte de cada variable independiente sobre \hat{y} . Otra forma de escribir la ecuación 2.132 es mediante un sistema de ecuaciones. Para el caso de dos variables independientes, x_1 y x_2 :

$$\beta_1 + r_{x_1x_2} \beta_2 = r_{x_1y} \quad (2.141)$$

$$r_{x_1x_2} \beta_1 + \beta_2 = r_{x_2y} \quad (2.142)$$

r_{x_1y} y r_{x_2y} son los coeficientes de regresión, r , de las rectas x_1y y x_2y , respectivamente, $r_{x_1x_2}$ se define como el coeficiente de correlación lineal simple (o de Pearson) entre dos variables x_1 y x_2 y se calcula según:

$$r(x_1, x_2) = \frac{S(x_1, x_2)}{S_{x_1} S_{x_2}} \quad (2.143)$$

donde S_{x_1}, S_{x_2} son las desviaciones estándar del conjunto de muestras 1 y 2, respectivamente y $S(x_1, x_2)$ es la covarianza muestral entre x_1 y x_2 y se calcula según:

$$S(x_1, x_2) = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (2.144)$$

y,

$$b_1 = \beta_1 \frac{S_y}{S_{x_1}} \quad (2.145)$$

$$b_2 = \beta_2 \frac{S_y}{S_{x_2}} \quad (2.146)$$

donde S_y, S_{x_1} y S_{x_2} son las desviaciones estándar totales de la población y, x_1 y x_2 , respectivamente.

Por último para calcular b_0 , se multiplican los coeficientes b_1 , b_2 calculados por los valores medios de x_1 y x_2 , respectivamente, se calcula el valor de \bar{y} :

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 \quad (2.147)$$

El modelo puede incluir efectos no lineales como x_1^2 , o efectos de interacción x_1x_2 .

Uno de los problemas no resueltos, que se pueden encontrar cuando se trata con RLM, son la colinealidad y la alta dimensionalidad de las variables x . Para solventar estos problemas, que pueden causar resultados no deseados o incluso la imposibilidad de ejecutar el modelo, es necesario tratar los datos originales para transformarlos en otros, aptos para el análisis de RLM [94]. Existen, principalmente, dos métodos: la regresión de componentes principales (PCR) y de mínimos cuadrados parciales (PLS).

La colinealidad, que se refiere a la dependencia que existe entre las variables predictoras, dificulta enormemente cuantificar el efecto que cada variable predictora tiene sobre la variable dependiente. Para evaluar la colinealidad de un conjunto de medidas existen dos parámetros denominados *factor de inflación de varianza* (Variance Inflation Factor, VIF) y el *número de condición*, η , que se calculan según las ecuaciones 2.148 y 2.149, respectivamente [95].

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2.148)$$

R_j^2 es el coeficiente de determinación obtenido al efectuar la regresión de X_j sobre el resto de las variables predictoras del modelo. Si $VIF \geq 10$ se considera que existe colinealidad.

El *número de condición*, η , es la prueba más adecuada para la detección de la colinealidad según Judge et al. [96]. Esta prueba se planteó inicialmente por Rachudel (1971) y se desarrolló años después por Belsley et al. [97] y se calcula según:

$$\eta = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \quad (2.149)$$

donde λ_{max} y λ_{min} son los autovalores máximos y mínimos, respectivamente de la matriz de datos estudiada. Cuando $\eta \geq 30$ se considera que sí existe colinealidad.

Como ya se ha mencionado PLS y PCR son técnicas que permiten solventar este problema y la alta dimensionalidad de las variables predictoras. En ambos casos, se transforman las variables en componentes ortogonales y, es necesario decidir cuáles son los que formaran parte de los modelos posteriores. Para ello, existen distintas formas de selección de componentes. Las más comunes son las explicadas en el apartado II.7.2.

II.8.2. Regresión en componentes principales

La PCR fue inicialmente propuesta por Massy [98] y trata de aplicar modelos de regresión lineal múltiple sobre un conjunto de componentes principales, obtenidos a partir de la matriz de correlación, tal y como se explica en la sección II.7.2. En el siguiente esquema de la figura II.9., se muestra la secuencia de pasos para construir un modelo PCR[94].

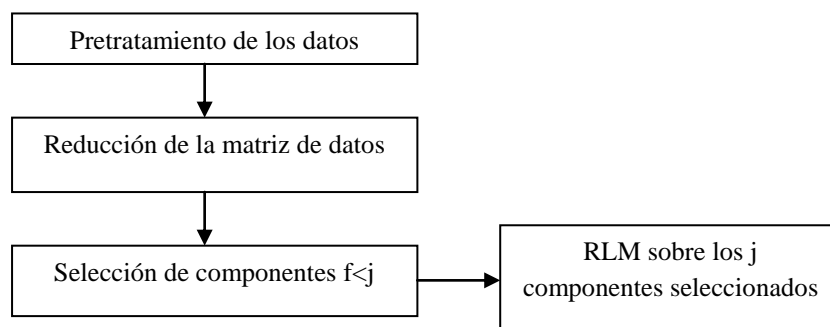


Figura II.9. Esquema de la secuencia de construcción de un modelo PCR.

II.8.3. Regresión por Mínimos Cuadrados Parciales

En el caso de PCR, se ha explicado cómo se usan los componentes principales del bloque x como variables predictoras por lo que las respuestas no tienen por qué estar

correlacionadas ya que sólo se emplea el bloque x . La PLSR, fue introducida por H.Wold [99], para aplicarse, sobre todo a ciencias sociales y económicas. Con los años ha ido ganando popularidad en el mundo de la quimiometría. El PLS es una técnica que, a día de hoy, sigue siendo objeto de estudio. La metodología PLS generaliza y combina características del PCA y RLM. En PLS se utilizan como variables predictoras vectores que explican menos varianza que los PCs pero que están más correlacionados con las respuestas. En PLSR suele haber dos pasos fundamentales, en primer lugar la transformación de la matriz de variables predictoras \mathbf{X} con ayuda del vector de respuestas \mathbf{Y} de orden $n \times 1$, en una matriz de componentes latentes no correlacionadas. Estos componentes son los que se denominan PLS. En segundo lugar, se calcula el modelo de regresión estimado usando el vector respuesta original como variables y como predictoras los componentes PLS seleccionados (apartado II.7.2)

Existen dos tipos de PLS en función de si se trata de una RLM o regresión lineal multivariante, PLS-1 y PLS-2, respectivamente. En este trabajo se utilizará en todos los casos PLS-1.

Las ventajas que proporcionan estos métodos frente a PCR son la reducción del impacto sobre el modelo de las variaciones del bloque X que no tienen correlación con la respuesta, y en el caso de PLS sí, y la reducción del número de componentes que explican la máxima varianza, ya que las direcciones del espacio que están poco relacionadas con las respuestas no se incluyen en el caso de PLS.

II.9. Reconocimiento de patrones mediante algoritmos basados en redes neuronales artificiales

Las redes neuronales consisten en algoritmos puramente matemáticos que trabajan mimetizando el funcionamiento del cerebro humano. El funcionamiento de estos métodos radica, principalmente, en cómo están interconectadas esas neuronas artificiales, es decir, en la arquitectura de la “red”. Precisamente, una de las razones por las que las redes neuronales gozan de una amplia adaptabilidad para el tratamiento de datos es la mencionada arquitectura de las redes, junto con el número de neuronas que las componen y la posibilidad de llevar a cabo tanto un aprendizaje supervisado como

no supervisado, por lo que tienen un amplio campo de aplicación. Dentro del campo de la química, existen numerosas aplicaciones con un elevado número de artículos publicados sobre control de procesos [100], predicción de la estructura secundaria de las proteínas [101], optimización de métodos cromatográficos [102] o a la clasificación de objetos en distintas categorías [103], entre otras.

Inicialmente, podemos considerar la red neuronal como una caja negra que acepta una serie de datos de entrada y proporciona uno o más datos de salida, tal y como se representa en la figura II.10.

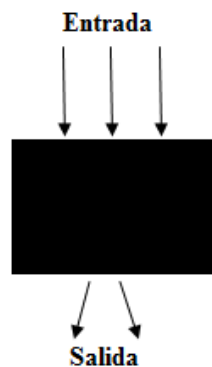


Figura II.10. Red neuronal en forma de caja negra.

En el interior de dicha caja existen unidades básicas conectadas unas con otras llamadas neuronas. Las entradas pasan a través de estas conexiones, que son las líneas de la red, transformándose y distribuyéndose en cada una de estas neuronas de forma que producen una o varias salidas [104]. Los datos de entrada son m -dimensionales, siendo la salida n -dimensional. Normalmente se utilizan redes en las que los datos de salida tienen una dimensión menor que los de entrada.

Las aplicaciones más comunes de las redes neuronales son [105]:

- *Autoasociación:* el sistema es capaz de reconstruir un dato de salida correcto a partir de un dato de entrada incompleto, es decir, identifica patrones corruptos.
- *Heteroasociación:* el sistema realiza asociación uno a uno entre los miembros de dos conjuntos de patrones.

- *Clasificación:* asigna a los objetos una clase en base a una o varias propiedades que caracterizan a una determinada clase. El proceso de clasificación puede realizarse tanto de forma supervisada como no supervisada. El entrenamiento se realiza con un subconjunto de los datos. Posteriormente, la red es capaz de predecir la clase o el cluster al cual pertenecen objetos desconocidos.
- *Transformación:* la red lleva a cabo la transformación de datos del espacio multivariante a otro tipo de datos, también multivariantes de igual o menor dimensionalidad manteniendo la topología de los datos, es decir, manteniendo la relación entre los datos.
- *Modelización:* consiste en buscar una función analítica o “modelo” que proporcione unos datos de salida n-dimensionales para una entrada m-dimensional. En la modelización clásica es necesario conocer de antemano la función de modelización; en el caso de los algoritmos neuronales sólo se necesita un número suficiente de datos de entrada, lo suficientemente espaciados como para que la red se adapte a cualquier relación no lineal entre los datos de entrada y los de salida.

II.9.1. Tipos y clasificaciones de las redes neuronales artificiales.

Existen una serie de modelos de ANN presentes en la bibliografía y en numerosos estudios. A lo largo de la historia se han propuesto diferentes clasificaciones para ordenar, de alguna manera, los tipos de redes existentes [106].

En primer lugar, existe una clasificación en función del patrón de conexiones que presentan, definiéndose tres tipos básicos de redes:

- *Redes de propagación hacia delante o acíclicas:* las señales van desde la capa de entrada a la de salida sin existir ciclos. En este grupo aparecen las redes monocapas (perceptrón o Adaline) y las multicapa (perceptrón multicapa)
- *Redes recurrentes:* presentan al menos un ciclo cerrado de activación neuronal (máquina de Boltzmann).

JJ. Montaña Moreno [107] expone en sus trabajos una clasificación que parte de los distintos sistemas de aprendizaje de las redes, y las va clasificando en función de los tipos de conexiones o las funciones de las mismas. Se trata de una clasificación exhaustiva que se resume en el esquema de la figura II.11.

- *Aprendizaje supervisado*: estas redes necesitan que los datos de entrada estén previamente clasificados o que el objetivo respuesta se conozca (perceptrón simple, Adaline, perceptrón multicapa, red de retropropagación (backpropagation), y la red de memoria asociativa bidireccional).
- *Aprendizaje no supervisado o autoorganizado*: no necesitan que se conozca una clasificación previa de los datos de entrada (redes de Hopfield, máquina de Boltzmann, redes de aprendizaje competitivo, o las redes de resonancia adaptativa).

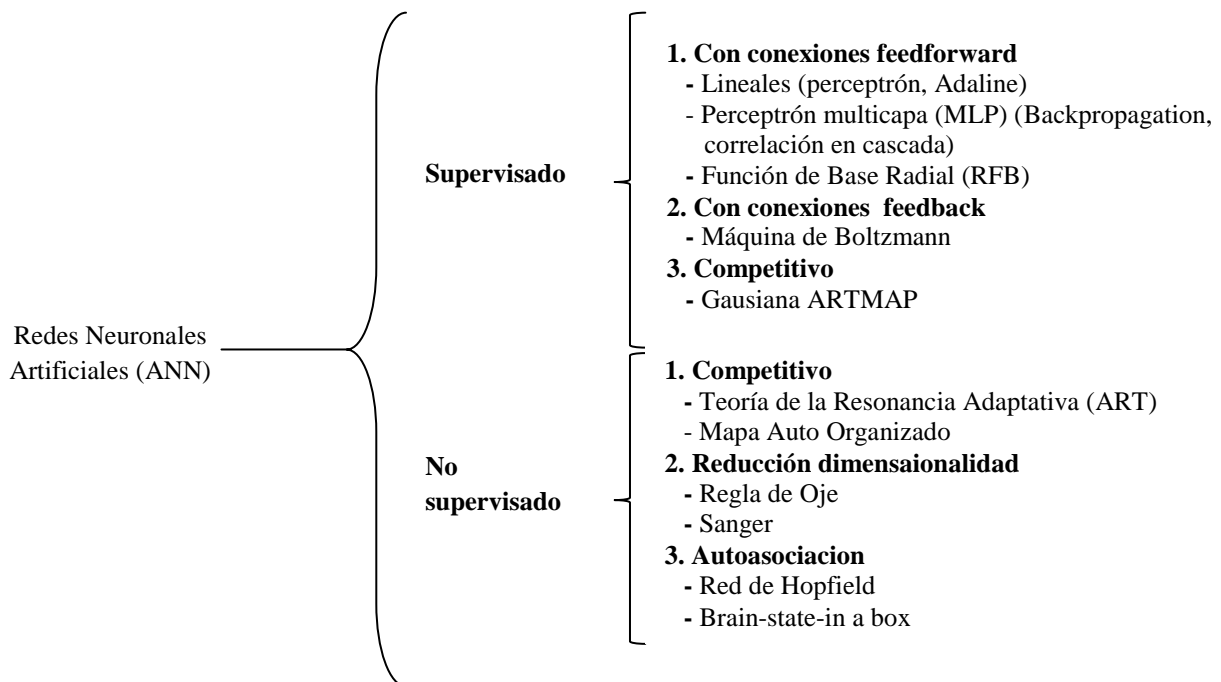


Figura II.11. Esquema propuesto para la clasificación de las ANN.

II.9.2. Neuronas y redes

Las redes neuronales artificiales son modelos matemáticos con una similitud muy superficial a las redes neuronales biológicas. En la neurobiología, a la modulación de la señal que se transmite entre las neuronas se denomina fuerza sináptica, y en el caso que nos ocupa, a esta fuerza sináptica se la denomina peso (w). Este peso va a determinar la cantidad de señal que entra en una neurona. Si asignamos la señal de entrada como $s_i (i = 1, \dots, m)$, que actúa sobre los pesos w_i , la señal de entrada global va a ser una función de todas las s_i que llegan en cada momento y de todos los pesos w_i , es decir, la señal global neta (Net) va a ser una suma de los productos $w_i s_i$. Se denomina función de propagación o de red (Net). En la ecuación siguiente se representa esta idea.

$$\begin{aligned}
 Net &= (Net_1, Net_2, \dots, Net_j, \dots, Net_n) = \\
 &= \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1m} \\ W_{21} & W_{22} & \dots & W_{2m} \\ \dots & \dots & \dots & \dots \\ W_{n1} & W_{n2} & \dots & W_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_i \\ \dots \\ x_{m-1} \\ 1 \end{bmatrix} \quad (2.150)
 \end{aligned}$$

Estas redes se van a someter a un proceso de aprendizaje para que puedan tomar decisiones que nos ayuden a nuestros propósitos de clasificación. Este proceso de aprendizaje en la red no es más que un cierto número de cambios en los valores de los pesos w_i . La expresión analítica de la señal Net es:

$$Net = w_1 s_1 + w_2 s_2 + \dots + w_m s_m \quad (2.151)$$

Las señales s_i se combinan en un vector de entrada X, cuyas componentes van a ser las s_i . De esta forma tenemos:

$$\{s_1, s_2, \dots, s_m\} = X(x_1, x_2, \dots, x_m) \quad (2.152)$$

De la misma forma las fuerzas sinápticas se escriben en forma de un vector de pesos W :

$$W = (w_1, w_2, \dots, w_m) \quad (2.153)$$

De manera que la función Net será que el producto escalar de los dos vectores W y X :

$$Net = WX \quad (2.154)$$

Generalmente se añade un off-set adicional llamado bias:

$$Net = WX + 0 \quad (2.155)$$

La adición de este término permite una mayor adaptabilidad para el aprendizaje. Por tanto, la expresión final de la función Net será:

$$Net = \sum_{i=1}^m w_i x_i + 0 \quad (2.156)$$

Todo proceso de aprendizaje implica el cálculo del vector de pesos óptimos que proporcione decisiones correctas. Para evaluar ese vector se utiliza la Regla Delta.

Inicialmente, el vector de pesos óptimo se escoge al azar y se calcula Net para un objeto determinado X . Si no se ha obtenido un resultado correcto, deberá modificarse W de la siguiente forma:

$$\Delta W = W^{(nuevo)} - W^{(viejo)} \quad (2.157)$$

Esta modificación es proporcional a una determinada cantidad δ del vector de pesos W :

$$\Delta W \approx \delta X \quad (2.158)$$

Por tanto, δ se puede expresar como:

$$\delta = \eta \left(\frac{\Delta W}{X} \right) \quad (2.159)$$

donde η es una constante de proporcionalidad llamada factor de aprendizaje. La ecuación estándar para la regla delta es:

$$\Delta W = \eta \delta X \quad (2.160)$$

Al principio del proceso de aprendizaje es conveniente mantener $\eta = 1$ y conforme va aumentando el tiempo de cálculo, mejorándose el aprendizaje, es aconsejable ir reduciendo el valor de η .

Una vez obtenido W y calculada Net , hay que transformar de forma no lineal esta función Net de forma análoga al comportamiento biológico. Esta modificación de Net para calcular la salida de la neurona se denomina *Función de transferencia* (Sal):

$$Sal = f(Net) \quad (2.161)$$

Esta función de transferencia debe cumplir tres requisitos:

- No ser negativa.
- Debe ser continua y estar confinada dentro de un determinado rango.
- Debe ser derivable.

Aunque existen muchas funciones matemáticas que cumplen estas condiciones, una de las más empleadas en aplicaciones de redes neuronales es la función sigmoidea, sf , o logística, que tiene la siguiente expresión:

$$sf(Net) = \frac{1}{1 + \exp(-Net)} \quad (2.162)$$

Una de las ventajas que presenta esta función de transferencia es que el cambio es el de máximo aprendizaje y una ventaja adicional es que la derivada de la función contiene a la propia función.

Existen otras funciones como la función identidad cuya expresión viene dada por la siguiente ecuación:

$$f(Net) = Net \quad (2.163)$$

la función exponencial o softmax o la función tangencial.

En la figura II.12. se presentan dos redes neuronales. En la parte superior, una red con una función tipo sigmoidea y en la parte inferior, otra con función identidad.

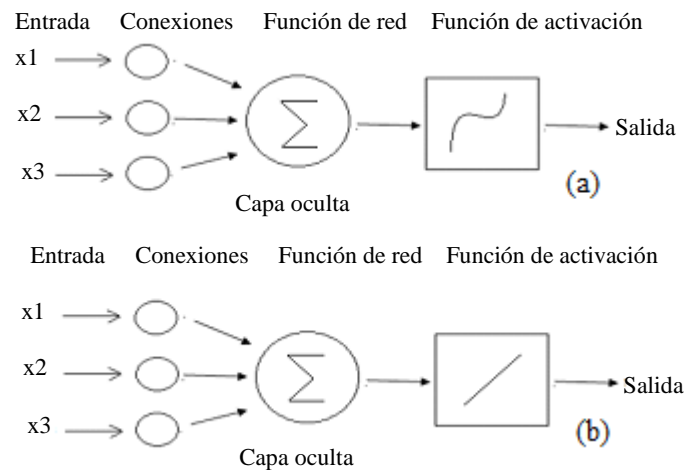


Figura II.12. Representación gráfica de una ANN con (a) función de activación sigmoidea; (b) ANN con función de activación identidad.

La señal de entrada debe adecuarse al conjunto de neuronas conectadas en la red, por lo tanto, la señal de entrada debe atacar a todas las neuronas a la vez, cada una de ellas con sus propios vectores de pesos, W , y señales de salida (Sal).

Al grupo de neuronas que reciben el mismo conjunto de señales de entrada X y producen otras tantas salidas simultáneas se les denomina *capa*. Todas las neuronas de la misma capa deben tener la misma dimensión. Como cada neurona j produce su propia función Net_j y su correspondiente señal de salida Sal_j , cada señal individual de una misma capa se puede englobar en los respectivos vectores \mathbf{x} , \mathbf{Net} y \mathbf{Sal} . El vector de salida \mathbf{Sal} de una capa será el vector de entrada \mathbf{X} de la siguiente capa de neuronas.

Dentro de la red se pueden distinguir tres tipos de capas[108]:

- *Capa de entrada*: es la capa que recibe directamente la información proveniente de las fuentes externas de la red, los datos de entrada. Es una capa no activa.
- *Capas ocultas*: son internas a la red y no tienen contacto directo con el entorno. El número de niveles ocultos puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras, lo que determina, junto con su número, las distintas topologías de redes neuronales.
- *Capa de salida*: es la que transfiere información de la red al exterior.

Independientemente de la arquitectura de la red, se debe cumplir que a todas las neuronas de la misma capa les ha de llegar igual número de entradas, incluyendo una entrada adicional (bias). Además, el número de pesos en cada neurona viene fijado por el número de señales procedentes de la capa anterior.

Desde el punto de vista matemático, vamos a considerar una capa de n neuronas, cada una de ellas con un vector de pesos m -dimensional, por lo que se tendrá una matriz de pesos W para una red multicapa. Cada elemento de esta matriz tendrá un superíndice indicativo de la capa a la cual pertenece y un subíndice j representativo de su neurona correspondiente (w_{ji}^l) .

Para aplicar una señal de entrada m -dimensional a una red neuronal con una capa de n neuronas, cada una de ellas con m pesos, tendremos que multiplicar un vector $\mathbf{X}(x_1, x_2, \dots, x_{m-1}, 1)$ por una matriz de pesos W de dimensión $n \times m$. El resultado es un vector $\mathbf{Net}(Net_1, Net_2, \dots, Net_m)$.

Para una red multicapa, se cumple que la señal de entrada de la capa l es la salida de la capa $l-1$:

$$X^l = Sal^{l-1} \quad (2.164)$$

Según esto se puede escribir:

$$Net_j^l = \sum_{i=1}^m w_{ij}^l sal_i^{l-1} \quad (2.165)$$

para $j=1,2,\dots,n$.

La señal de salida de la capa l se obtiene a partir de Net^l mediante la señal de transferencia, en el caso de utilizar la función sigmoidea:

$$Sal^l = sf(Net^l) \quad (2.166)$$

II.9.3. Aprendizaje no supervisado

En el caso de algoritmos neuronales con aprendizaje no supervisado la red debe descubrir por sí sola características, correlaciones o categorías de los datos de entrada y obtenerlas de forma codificada a la salida. En este sentido se puede decir que estas unidades y conexiones muestran cierto grado de autoorganización. Uno de los modelos más característicos de redes no supervisadas es el de mapas autoorganizados de Kohonen [109].

II.9.4. Aprendizaje supervisado

En estos sistemas se dispone de algún tipo de información que permite decidir cuándo dejar de aprender, con qué intensidad aprender o cada cuánto tiempo hacerlo. Rumelhart et al. [110] idearon un método, basado en otros investigadores [111,112], para que una red neuronal aprendiera la asociación existente entre los patrones de entrada a la misma y las clases correspondientes empleando varias capas de neuronas. Este método es el que se conoce como *backpropagation* error de retropropagación.

La principal ventaja de este tipo de aprendizaje es que la corrección de los pesos está muy bien definida. Dicha corrección empieza a ser aplicada a los pesos de la última capa (capa de salida) continuando hacia atrás hasta llegar a la capa de entrada. En la figura II.13. se muestra un esquema de la secuencia que se sigue cuando se tiene una red de aprendizaje mediante la corrección *backpropagation*.

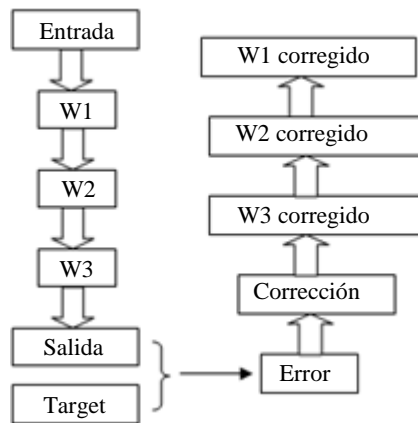


Figura II.13. Corrección de pesos en el aprendizaje por retropropagación.

La arquitectura de la red de retropropagación es muy flexible, de forma que el número de capas así como el número de neuronas que las forman varían según la aplicación para la cual se está desarrollando la red. En la mayoría de los casos una red de retropropagación consiste en una capa de entrada con tantas neuronas como componentes tenga la señal de entrada, y dos capas activas: la oculta y la capa de salida.

El proceso de aprendizaje es supervisado porque precisa de un conjunto predefinido de parejas de datos de entrada, que consiste en objetos (X) junto con unos objetivos o targets asociados (T). Durante el proceso de entrenamiento, se ajustan los pesos de la red con el fin de que las entradas introducidas produzcan las salidas deseadas. Los pasos a seguir en el entrenamiento son:

- a) Elegir dos conjuntos de objetos de entrada y de objetivos e introducir el primer vector de entrada (X) en la red.
- b) Calcular la salida de la red.
- c) Calcular el error entre la salida de la red y la salida deseada (Target).
- d) Modificar los pesos de la red con el fin de minimizar el error, empleando la regla Delta.

Se repiten los pasos anteriores para cada vector del conjunto de entrenamiento hasta disminuir convenientemente el error. Las neuronas de la capa de entrada actúan como *buffers* de entrada de los datos, los cuales son ponderados por cada una de las neuronas de la capa. La siguiente capa, es decir, la capa oculta, recibe las salidas ya ponderadas de la capa de entrada, suma todas las entradas en cada neurona y pasa dicha

suma a través de la función de transferencia. El resultado obtenido pondera y pasa hacia la capa de salida donde se repite el proceso de suma y transformación. El valor de la función de transferencia de la capa de salida constituye la respuesta de la red del objeto introducido.

El proceso de entrenamiento se realiza de forma iterativa introduciendo en la red todos los miembros del conjunto de parejas conocido objeto-objeto (set entrenamiento). Cada paso del conjunto completo se denomina *época*. El aprendizaje termina cuando no se mejore la diferencia entre la salida de la red y el objetivo correspondiente, o bien cuando se llegue a un número determinado de épocas.

II.10. Bibliografía

1. C. M. Cuadras. *Nuevos Métodos de Análisis Multivariante*, CMC, Barcelona, 2014.
2. Grupo de Quimiometría y Cualimetría de Tarragona. *Quimiometría, una disciplina útil para el análisis químico*. www.quimica.urv.net/quimio/quimio.pdf
3. D.L. Massart. *Handbook of Chemometrics and Qualimetrics*. Elsevier, Amsterdam, 1997.
4. S. Wold, M. Sjostrom. SIMCA: a method for analyzing chemical data in terms of similarity analogy. *The ACS Symposium Series*, **1977**, 52, 243-282.
5. M. Bina, P. Bae, P. Gaillard. Pattern recognition machine: its conception and creation. *Microprocessing and Microprogramming*, **1987**, 19, 327-333.
6. G.R. Ramos, C. García-Álvarez Coque. *Quimiometría*. Síntesis, Madrid, 2001.
7. J.F. Hair, R.E. Anderson, R.L. Tatham. *Análisis Multivariante de Datos*, 5ªed., Prentice Hall, Madrid, 1999.
8. N. González, J. Abad, J.P. Lèvy. Normalidad y otros supuestos en análisis de covarianzas. *Modelización con Estructuras de Covarianzas*, 2006, Netbiblo, La Coruña, 31-57.
9. S. Sharma. *Applied Multivariate Techniques*. Wiley & Sons. Nueva York., 1996
10. J.L. Bonnet Pérez. *Lecciones de estadística descriptiva y probabilidad*. Club Universitario. Alicante, 2004.
11. S. S. Shapiro, M. B. Wilk. An analysis of variance test for normality (Complete Samples). *Biometrika*, **1965**, 52, 591-611.
12. M.A. Stephens. EDF Statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, **1974**, 69, 730-737.
13. H.W. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, **1967**, 62, 399-402.
14. E. S. Pearson. Student as statistician, *Biometrika*, **1939**, 30, 210-250.
15. R. Romero Villafranca, L.R. Zúnica Ramajo. *Métodos Estadísticos en Ingeniería*. Universidad Politécnica de Valencia, 2005.
16. M. F. Triola. *Estadística*. Pearson Educación, México, 2004.
17. D. W. Zimmerman. Teacher's corner: A note on interpretation of the paired-samples *t*-test. *Journal of Educational and Behavioral Statistics*, **1997**, 22, 349-360.

18. B. L. Welch. The generalization of 'Student's' problem when several different population variances are involved, *Biometrika*, **1947**, 34, 28-35.
19. R. Compañó, A. Ríos. *Garantía de la Calidad en los Laboratorios Analíticos*. Editorial Síntesis, Madrid, 2002.
20. J. M. Jurado. *Estadística Básica Aplicada al Laboratorio Analítico. Calibración Mediante Regresión Lineal*. Programa de formación del CSIC: Control de calidad en Laboratorios de Análisis Químicos, 2015.
21. J.N. Miller, J.C. Miller. *Statistics and Chemometrics for Analytical Chemistry*. Prentice Hall, Essex, 2005.
22. C. Guisande González, A. Vaamonde Lise, A. Barreiro Felpeto. *Tratamiento de datos con R, Statistica y SPSS*. Díaz de Santos, Madrid, 2013.
23. A. Díaz. *Diseño Estadístico de Experimentos*, 2ª Ed. Universidad de Antioquía, 2009.
24. D. C. LeBlanc. *Statistics: Concepts and Applications for Science*. Jones & Bartlett Learning, New York, 2004
25. J.A. Rafter, M.L. Abell, J. P. Braselton. Multiple comparison methods for means. *SIAM Review*, **2002**, 44, 259-278.
26. H.L. Harter. Tables of range and studentized range. *Annals of Mathematical Statistics*, **1960**, 31, 1122-1147.
27. C. Mongay. *Quimiometría*. Universidad de Valencia, 2011.
28. H. B. Mann, D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, **1947**, 22, 125-128.
29. L. Salafranca i Cosialls, A. Solana Pérez, M.I. Núñez Peña, M. Jiménez Fernández, D. Miralles Pinta, G. Serra Delgado. *Estadística aplicada con SPSS y Statgraphics*. Universitat Barcelona, 2000.
30. J. E. de Muth. *Basic Statistics and Pharmaceutical Statistical Applications*, CRC Press, Boca Ratón, 2014.
31. W. H. Kruskal, W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, **1959**, 47, 583-621.
32. A. Vargas Sabadías. *Estadística Descriptiva y Diferencial*. Universidad de Castilla-La Mancha, 1996.

33. R. Delgado de la Torre. *Probabilidad y Estadística para Ciencias e Ingenierías*. Delta Publicaciones, Madrid, 2007.
34. ISO 3534. *Statistics. Vocabulary and Symbols. Part 1: Probability and General Statistical Terms, section 2.64.*, Geneva, 1993.
35. S. Burke. Missing values, outliers, robust & non-parametric methods. *LC-GC Europe Online Supplement*, **1999**, 19-24.
36. AOAC Official Methods Program Manual, Part 6: Guidelines for Collaborative Study. http://www.aoac.org/imis15_prod/AOAC_Docs/StandardsDevelopment/Collaborative_Study_Validation_Guidelines.pdf
37. F. Grubbs. Procedures for detecting outlying observations in samples. *Technometrics*, **1969**, 11, 1-21.
38. P. Mair, A. von Eye. *Cochran's test. Encyclopedia of Statistical Behavioral Science*. Wiley & Sons, New York, 2005.
39. S. L. R. Ellison, T. J. Farrant, V. Barwick. *Practical Statistics for the Analytical Scientist: A Bench Guide*, Royal Society of Chemistry, London, 2009.
40. F. Yates. *The Design and Analysis on Factorial Experiments*, Farnham Royal Commonwealth Agriculture Bureaux, 1959.
41. E. Morgan, K.W. Burton, P.A. Church. Practical exploratory designs. *Chemometrics and Intelligent Laboratory System*, **1989**, 5, 283-302.
42. E. del Castillo. Modified desirability functions for multiple response optimization. *Journal of Quality Technology*, **1996**, 28, 337-345.
43. A. G. González. Two levels factorial designs base on multiple linear regression models: a tutorial digest illustrated by case studied. *Analytica Chimica Acta*, **1998**, 360, 227-241.
44. A. Delacroix, C. Porte. Utilisation de plans d'expériences associés à des methods directes pour l'optimisation en chimiométrie. *Analisis*, **1996**, 24, M22-M25.
45. G.T. Wernimont. *Use of Statistics to Develop and Evaluate Analytical Methods*. Ed., William Spendley, AOAC, Virginia, 1990.
46. I. Sánchez. *Métodos Estadísticos para la Mejora de la Calidad*. Universidad Carlos III, Madrid.
http://portal.uc3m.es/portal/page/portal/dpto_estadistica/personal/ismael_sanchez_rodri_guez_morcillo.

47. J. Harrington. The desirability function. *Industrial Quality Control*, **1965**, 21, 494-498.
48. G.C. Derringer, D. Suich. Simultaneous optimization of several response variables. *Journal of Quality Technology*, **1980**, 12, 214-219.
49. M. Jimidar, B. Bourguignon, D. L. Massart. Application of Derrington's desirability function for the selection of optimum separation conditions in capillary zone electrophoresis. *Journal of Chromatography A*, **1996**, 740, 109-117.
50. B. Bourguignon, D.L. Massart. Simultaneous optimization of several chromatographic performance goals using Derringer's desirability function. *Journal of Chromatography A*, **1991**, 586, 11-20
51. Statsoft, STATISTICA for Windows, Computer Program Manual, Tulsa, OK, **1999**. <http://www.statsoft.com>.
52. B.R. Kowalski, C.F. Bender. Pattern recognition machine: its conception and creation. *Journal of American Chemical Society*, **1972**, 94, 5632-5639.
53. C. Albano, W.J. Dunn III, E. Edlund, E. Johanson, B. Norden, M. Sjöström, S. Wold. Four levels of pattern recognition. *Analytica Chimica Acta*, **1978**, 103, 429-433.
54. S.W. Page. Pattern Recognition methods for the determination of food composition. *Food Technology*, **1986**, 40, 104-109.
55. K. Varmuza, *Pattern Recognition in Chemistry*. Springer-Verlag. Heidelberg, 1980, 25-29.
56. R. Fisher. Inverse probability and the use of likelihood. *Proceedings of the Cambridge Philosophical Society*, **1932**, 28, 257-261
57. M. Meloun, J. Militky, M. Forina. *Chemometrics for Analytical Chemistry. Vol I: PC-Aided Statistical Data Analysis*. Ellis Horwood Limited, Chichester, 1992.
58. B.R. Kowalski, C.F. Bender. Pattern Recognition: a powerful approach to interpreting chemical data. *Journal of the American Chemical Society*, **1972**, 94, 5632-5639.
59. R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annual Eugenics*, **1936**, 7, 179-188.
60. D. Coomans, M. Jonckheer, D. L. Massart, I. Broeckert, P. Blockx. The application of linear discriminant analysis in the diagnosis of thyroid diseases. *Analytica Chimica Acta*, **1978**, 103, 409-415.

61. S. Wilks. Multidimensional scatter in Olkin. *Contributions to Probability and Statistics*, Standford Press, **1960**, 597-614.
62. G.P. McCabe. *Computations for Variable Selection of Analytical Data. Encyclopedia of Analytical Chemistry*. John Willey & Sons, Chichester, 2000.
63. I. T. Jolliffe. *Principal Component Analysis*. Springer Science & Business Media, New York, 2002.
64. B. Chanda, D. Dutta Majumder. *Digital Image Processing and Analysis*. PHI Learning Pvt., 2004.
65. H.F. Kaiser. The application of electronic computers in factor analysis. *Educational and Psychological Measurements*, **1960**, 20, 141-151.
66. J.R. Piggott. *Statistical Procedures in Food Research*. Elsevier Applied Science. London, 1996.
67. J. Shao. Linear model selection by cross-validation. *Journal of the American Statistical Association*, **1993**, 88, 486- 494.
68. M. Blanco, V. Cerdá. *Temas Avanzados de Quimiometría*. Universitat Illes Balears, 2007.
69. K.R. Gabriel. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **1971**, 58, 453-467.
70. H.C. Romesburg. *Cluster Analysis for Researchers*. Lifetime Learning Publications, Belmont, California, 1984.
71. K. Varnuza. Pattern recognition in analytical chemistry. *Analytica Chimica Acta*, **1980**, 122,227-240.
72. L. Kryger. Interpretation of analytical chemical information by pattern recognition methods: a survey. *Talanta*, **1981**, 28, 871-887.
73. J.L. Bienias, P.S: Kott, D.A. Evans. *Applying the delete-a-group jackknife variance estimator to analyses of data from a complex longitudinal survey*. Proceedings of the Annual Meeting of the American Statistical Association, Section on Survey. American Statistical Association. 2003.
74. P. Refaelzadeh, L. Tang, H. Lui. *Cross Validation*. Arizona State University, 2008.
75. A. J. Stuper, P.C. Jurs. Structure activity studies of barbiturates using pattern recognition techniques. *Journal of Pharmaceutical Science*, 1978, 67, 745-751.

76. O. VladimirovnaPanteleeva. *Fundamentos de Probabilidad y Estadística*. UAEM, México, 2005.
77. L. F. Escudero. *Reconocimiento de patrones*. Paraninfo, Madrid, 1977.
78. R.A. Fisher. The statistical utilization of multiple measurements. *Annals of Eugenics*, **1938**, 8, 376-378.
79. J.J. Powers, E.S. Keith. Stepwise discriminant analysis of gas chromatography data as an aid in classifying the flavor quality of foods. *Journal of Food Science*, **1968**, 33, 207-213.
80. C.R. Rao. *Linear Statistical Inference and its Applications*. Wiley, New York, 1965.
81. D. Coomans, D.L. Massart, L. Kaufman. Optimization by statistical linear discriminant analysis in analytical chemistry. *Analytica Chimica Acta*, **1979**, 112, 97-122.
82. C. Cortes, V. Vapnik. Support vector machines for classification and regression. *Analyst*, **2009**, 135, 230-267.
83. W. M. Campbell, J. P. Campbell, A. Reynolds,, E. Singer, y P. A. Torres Carrasquillo. Support vector machines for speaker and language recognition. *Computer Speech and Language*, **2006**,20, 210-229.
84. E. Romero, L. Márquez, X. Carreras. Margin maximization with feed-forward neural networks: a comparative study with SVM and AdaBoost. *Neurocomputing*, **2004**, 57, 313-344.
85. N. Chen, W. Lu, J. Yang, G. Li. *Support Vector Machine*. World Scientific, Singapore, 2004.
86. A. Ben-Hur, C.S. Ong, S. Sonnenburg, B. Schölkopf, G. Rätsch. Support Vector Machines and Kernels for Computational Biology. *PLOS Computational Biology*, **2008**, 4, 1-10.
87. S. Wold. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics*, **1978**, 58, 453-467.
88. S. Wold, M. Sjöström. *SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy*. ACS Symposium Series. Data Processing Congress; B.R. Kowalski (Ed.), Chapter 12, 243-282, 1997.

89. S. Wold, K. Anderson. Major components influencing retention indices in gas chromatography. *Journal of Chromatography*, **1973**, 80, 43-59.
90. S. Wold, M. Sjöström. *Linear Free Relationships as Tools for Investigating Chemical Similarity-Theory as practice, in Correlation Analysis in Chemistry*. Ed. N.B. Chapman, J. Shorter, Plenum, Nueva York, 1978.
91. X.Liu, P. van Espen, F. Adams, S.H. Yan, M. Vanbelle. Classification of Chinese tea samples according to origin and quality by principal component technique. *Analytica Chimica Acta*, **1987**, 200, 421-430.
92. R. E. Walpole, R. H. Myers, S. L. Myers. *Probabilidad y Estadística para Ingenieros*. Pearson Educación, Madrid, 1999.
93. M.C. Persico Jiménez. *Regresión Lineal Múltiple*, Universidad Central Chile, Facultad de Ciencias Económicas y Administrativas, Santiago de Chile.
94. J.C. Vega-Vilca, J. Guzmán. Regresión en PLS y PCA como solución al problema de multicolinealidad en regresión múltiple. *Revista de Matemática: Teoría y Aplicaciones*, **2011**, 18, 9-20.
95. R. Álvarez Cáceres. *Estadística Aplicada a las Ciencias de la Salud*. Díaz de Santos, Madrid, 2007.
96. G.G. Judge, R.C. Hill, W. Griffiths, H. Lutkepohl, T.C. Lee. *Introduction to the Theory and Practice of Econometrics*. John Wiley & Sons, California, 1985.
97. D.A. Belsley, E. Kuh, R.E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, Nueva York, 1980.
98. W.F. Massy. Principal Components Regression in Exploratory Statistical Research. *Journal of the American Statistical Association*, **1965**, 60, 234–246.
99. H. Wold. Soft modeling by latent variables; the nonlinear iterative partial least square approach. *Perspectives in Probability and Statistics*, **1975**, Papers in Honour of M.S. Bartlett.
100. G. Asadollahfardi. *Water Quality Management: Assessment and Interpretation*. Springer, Berlin Heidelberg, 2015.
101. N. Qian, T.J. Sejnowski. Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology*, **1988**, 202, 865–884.
102. A. A. D'Archivo, M.A. Maggi, F. Ruggieri. Artificial neural network prediction of multilinear gradient retention in reversed-phase HPLC: comprehensive QSRR-based

models combining categorical or structural solute descriptors and gradient profile parameters. *Analytical and Bioanalytical Chemistry*, **2015**, 407, 181-190.

103. J. Khan, J. S. Wei, M. Ringnér, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, P. S. Meltzer. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, **2001**, 7, 673 – 679.

104. S. A. Kalogirou. Artificial neural networks in renewable energy systems applications: a review. *Renewable and Sustainable Energy Review*, **2001**, 5, 373–401.

105. J. Zupan, J. Gasteiger. *Neural Networks for Chemists: An Introduction*. John Wiley & Sons, New York, 1993.

106. K. Swingler. *Applying Neural Networks: A Practical Guide*. Paperback, Morgan Kaufmann, 1996.

107. J. J. Montaña Moreno. *Redes Neuronales Artificiales Aplicadas a Análisis de Datos*. Tesis doctoral, Universitat de les Illes Balears, 2002. <http://www.tdx.cat/bitstream/handle/10803/9441/tjjmm1de1.pdf?sequence=1>.

108. J. de Villers, E. Bernard. Backpropagation neural nets with one and two hidden layers. *Neural Networks, IEEE Transactions*, **1993**, 4, 136-141.

109. T. Kohonen, A. Mäkisara, O. Simula, J. Kangas. *The Hypermap Architecture. Artificial Neural Network*, Elsevier Science Publishers, Amsterdam, 1991.

110. D.E. Rumelhart, J.L. McClelland. An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, **1982**, 89, 60-94.

111. P. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Ph. Thesis, Harvard University, 1974.

112. D. Parker. *Learning logic*. Invention Report, 581-664. Office of Technology Licensing, Stanford University, 1982.

III. Parte Experimental

III.1. Material y reactivos

III.1.1. Aparatos

- Espectrómetro de plasma inducido acoplado ULTIMA 2 (HORIBA JOBIN YVON, Kioto, Japón) equipado con:
 - Antorcha Minitorch.
 - Nebulizador Meinhard.
- Cromatógrafo líquido (HITACHI, Tokyo, Japón):
 - Bomba binaria para HPLC Hitachi Elite L-2130.
 - Inyector Rheodyne con bucle de 20 µl.
 - Detector UV Hitachi Elite L-2400.
 - Columna cromatográfica TRACER EXCEL de fase inversa, C18, 150 mm de longitud, 4 mm d.i., 5 µm de tamaño de partícula.
 - Horno para columna GECKO 2000.
- Espectrofotómetro Thermo Spectronic Unicam UV 500
- Baño de arena P Selecta Combiplac.
- Baño de Ultrasonido P Selecta Ultrasons H-D.
- Estufa Memmert UNB 500 (Schwabach, Alemania)

III.1.2. Reactivos y disoluciones

III.1.2.1. Reactivos

Todos los reactivos utilizados en la realización de este trabajo fueron de, al menos, calidad para análisis.

- Acetona, CH₃COCH₃, (Panreac, Barcelona, España)
- Acetonitrilo, CH₃CN, calidad gradiente para cromatografía de líquidos (Teknokroma, Barcelona, España)
- Ácido bórico, H₃BO₃, (Merck, Darmstadt, Alemania)
- Ácido clorhídrico, HCl, (37%) (Merck, Darmstadt, Alemania)

- Ácido nítrico, HNO_3 , (65%) (Merck, Darmstadt, Alemania)
- Ácido perclórico, HClO_4 , (60%) (Merck, Darmstadt, Alemania)
- Ácido sulfúrico, H_2SO_4 , (96%) (Merck, Darmstadt, Alemania)
- Amonio dihidrogeno fosfato, $(\text{NH}_4)\text{H}_2\text{PO}_4$, (Merck, Darmstadt, Alemania)
- Capsaicina, $\text{C}_{17}\text{H}_{27}\text{NO}_3$, (Sigma Aldrich, Saint Louis, USA)
- Carbonato de calcio, CaCO_3 , (Sigma Aldrich, Saint Louis, USA)
- Cloruro de manganeso, MnCl_2 , (Sigma Aldrich, Saint, Louis, USA).
- Cloruro de níquel, NiCl_2 , (Sigma Aldrich, Saint Louis, USA)
- Cloruro de potasio, KCl , (Sigma Aldrich, Saint Louis, USA)
- Cloruro de sodio, NaCl , (Sigma Aldrich, Siant Louis, USA)
- Cobre metálico, Cu , (Merck, Darmstadt, Alemania)
- Dicromato de potasio, $\text{K}_2\text{Cr}_2\text{O}_7$, (Merck, Darmstadt, Alemania).
- Dihidrocapsaicina, $\text{C}_{17}\text{H}_{29}\text{NO}_3$, (Sigma Aldrich, Saint Louis, USA)
- Hierro metálico, Fe , (Merck, Darmstadt, Alemania)
- Peróxido de hidrógeno, H_2O_2 , (30%) (Merck, Darmstadt, Alemania)
- Metanol, CH_3OH , calidad gradiente para cromatografía de líquidos, (Teknokroma, Barcelona, España)
- Nitrato de aluminio, $\text{Al}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$, (Merck, Darmstadt, Alemania)
- Nitrato de cobalto, $\text{Co}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$, (Merck, Darmstadt, Alemania)
- Nitrato de magnesio, $\text{Mg}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$, (Sigma Aldrich, Saint Louis, USA)
- Nitrato de plomo, $\text{Pb}(\text{NO}_3)_2$, (Merck, Darmstadt, Alemania).
- Nitrato de estroncio, $\text{Sr}(\text{NO}_3)_2$, (Sigma Aldrich, Saint Louis, USA).
- Sulfato de amonio, $(\text{NH}_4)_2\text{SO}_4$, (Merck, Darmstadt, Alemania).
- Sulfato de cinc, $\text{ZnSO}_4 \cdot \text{H}_2\text{O}$, (Merck, Darmstadt, Alemania).

III.1.2.2. Disoluciones estándar

- Las disoluciones estándar (1000 mg L^{-1}) de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr, Zn, se prepararon de acuerdo con lo descrito en la guía de preparación de patrones de espectroscopía atómica de PERKIN-ELMER (Norwalk, CN) [1].

- Las disoluciones de trabajo se obtuvieron por dilución de las anteriores empleando siempre agua tratada con un sistema Milli-Q (Millipore, Bedford, MA) con una resistividad de 18,2 M Ω -cm.
- Capsaicina (250 mg L⁻¹). Se toman 125 mg, medidos con precisión hasta la décima de mg, de reactivo puro enrasando con 500 mL de metanol.
- Dihidrocapsaicina (250 mg L⁻¹). Se toman 125 mg, medidos con precisión hasta la décima de mg, de reactivo puro enrasando con 500 mL de metanol.
- Las disoluciones de trabajo de capsaicina y dihidrocapsaicina se obtuvieron por dilución de las anteriores empleando siempre agua Milli-Q.
- Disolución de sulfato de amonio y cobalto (II) hexahidratado (34,96 g L⁻¹) y dicromato potásico (0,3005 g L⁻¹) en disolución de ácido sulfúrico 1,8 M.
- Disolución de sulfato de amonio y nitrato de cobalto (II) (para 100 mL): 10 mL de sulfúrico concentrado 98%, 3,54 g de nitrato de cobalto (II) y 0,03 g de sulfato de amonio.

III.1.3. Muestras

Se han analizado 144 muestras de pimentón de las denominaciones de origen Murcia y Extremadura [2,3], de diversas marcas comerciales adquiridas en comercio minorista. 72 son de la DOP Extremadura y 72 de la DOP Murcia. En la tabla III.1 se incluye un listado de las muestras estudiadas con indicación del código utilizado para su identificación, la marca del producto, así como el origen geográfico y el tipo de pimentón, dulce, picante o agridulce. El código de cada muestra incluye una referencia de la DOP correspondiente, un número de orden y un superíndice que indica si se trata de pimentón dulce (d), picante (p) o agridulce (a).

Tabla III.1. Muestras de pimentón estudiadas

Código	Origen	Marca	Tipo
EX01 ^a	Extremadura	Tiburcio Pérez Trancón	Agridulce
EX02 ^a	Extremadura	Tiburcio Pérez Trancón	Agridulce
EX03 ^a	Extremadura	La Chinata	Agridulce
EX04 ^a	Extremadura	La Chinata	Agridulce
EX05 ^a	Extremadura	Ntra. Sra. Del Salobral	Agridulce
EX06 ^a	Extremadura	Ntra. Sra. Del Salobral	Agridulce
EX07 ^a	Extremadura	Los Extremeños	Agridulce
EX08 ^a	Extremadura	Los Extremeños	Agridulce
EX09 ^a	Extremadura	Vega 31 Cáceres	Agridulce
EX10 ^a	Extremadura	Vega 31 Cáceres	Agridulce
EX11 ^a	Extremadura	Carlos I	Agridulce
EX12 ^a	Extremadura	Carlos I	Agridulce
EX13 ^a	Extremadura	El rey de la Vera	Agridulce
EX14 ^a	Extremadura	El rey de la Vera	Agridulce
EX15 ^a	Extremadura	El rey de la Vera	Agridulce
EX16 ^a	Extremadura	El rey de la Vera	Agridulce
EX17 ^a	Extremadura	El Colorín	Agridulce
EX18 ^a	Extremadura	El Colorín	Agridulce
EX19 ^a	Extremadura	Jaseret	Agridulce
EX20 ^a	Extremadura	Jaseret	Agridulce
EX21 ^a	Extremadura	El caballo de oros	Agridulce
EX22 ^a	Extremadura	El caballo de oros	Agridulce
EX23 ^a	Extremadura	Paz	Agridulce
EX24 ^a	Extremadura	Paz	Agridulce
EX25 ^d	Extremadura	La Chinata	Dulce
EX26 ^d	Extremadura	La Chinata	Dulce
EX27 ^d	Extremadura	Tiburcio Pérez Trancón	Dulce
EX28 ^d	Extremadura	Tiburcio Pérez Trancón	Dulce
EX29 ^d	Extremadura	La Chinata	Dulce
EX30 ^d	Extremadura	La Chinata	Dulce
EX31 ^d	Extremadura	Ntra. Sra. Del Salobral	Dulce
EX32 ^d	Extremadura	Ntra. Sra. Del Salobral	Dulce
EX33 ^d	Extremadura	Los Extremeños	Dulce
EX34 ^d	Extremadura	Los Extremeños	Dulce
EX35 ^d	Extremadura	Vega 31 Cáceres	Dulce
EX36 ^d	Extremadura	Vega 31 Cáceres	Dulce
EX37 ^d	Extremadura	Carlos I	Dulce
EX38 ^d	Extremadura	Carlos I	Dulce

Tabla III.1. Continuación.

Código	Origen	Marca	Tipo
EX39 ^d	Extremadura	El rey de la Vera	Dulce
EX40 ^d	Extremadura	El rey de la Vera	Dulce
EX41 ^d	Extremadura	El rey de la Vera	Dulce
EX42 ^d	Extremadura	El rey de la Vera	Dulce
EX43 ^d	Extremadura	El Colorín	Dulce
EX44 ^d	Extremadura	El Colorín	Dulce
EX45 ^d	Extremadura	Gualtaminos	Dulce
EX46 ^d	Extremadura	Gualtaminos	Dulce
EX47 ^d	Extremadura	SPP	Dulce
EX48 ^d	Extremadura	SPP	Dulce
EX49 ^p	Extremadura	La Chinata	Picante
EX50 ^p	Extremadura	La Chinata	Picante
EX51 ^p	Extremadura	Triburcio Pérez Trancón	Picante
EX52 ^p	Extremadura	Triburcio Pérez Trancón	Picante
EX53 ^p	Extremadura	La Chinata	Picante
EX54 ^p	Extremadura	La Chinata	Picante
EX55 ^p	Extremadura	Ntra. Sra. Del Salobral	Picante
EX56 ^p	Extremadura	Ntra. Sra. Del Salobral	Picante
EX57 ^p	Extremadura	Los Extremeños	Picante
EX58 ^p	Extremadura	Los Extremeños	Picante
EX59 ^p	Extremadura	Vega 31 Cáceres	Picante
EX60 ^p	Extremadura	Vega 31 Cáceres	Picante
EX61 ^p	Extremadura	Carlos I	Picante
EX62 ^p	Extremadura	Carlos I	Picante
EX63 ^p	Extremadura	El rey de la Vera	Picante
EX64 ^p	Extremadura	El rey de la Vera	Picante
EX65 ^p	Extremadura	El rey de la Vera	Picante
EX66 ^p	Extremadura	El rey de la Vera	Picante
EX67 ^p	Extremadura	El Colorín	Picante
EX68 ^p	Extremadura	El Colorín	Picante
EX69 ^p	Extremadura	Paz	Picante
EX70 ^p	Extremadura	Paz	Picante
EX71 ^p	Extremadura	SPP	Picante
EX72 ^p	Extremadura	SPP	Picante
MU01 ^a	Murcia	Oro molido	Agridulce
MU02 ^a	Murcia	Oro molido	Agridulce
MU03 ^a	Murcia	Oro molido	Agridulce
MU04 ^a	Murcia	Oro molido	Agridulce
MU05 ^a	Murcia	Oro molido	Agridulce

Tabla III.1. Continuación.

Código	Origen	Marca	Tipo
MU06 ^a	Murcia	Oro molido	Agridulce
MU07 ^a	Murcia	Exportación Cervantes	Agridulce
MU08 ^a	Murcia	Exportación Cervantes	Agridulce
MU09 ^a	Murcia	Exportación Cervantes	Agridulce
MU10 ^a	Murcia	Oro molido	Agridulce
MU11 ^a	Murcia	Oro molido	Agridulce
MU12 ^a	Murcia	Oro molido	Agridulce
MU13 ^a	Murcia	Oro molido	Agridulce
MU14 ^a	Murcia	Oro molido	Agridulce
MU15 ^a	Murcia	Oro molido	Agridulce
MU16 ^a	Murcia	Ecoato	Agridulce
MU17 ^a	Murcia	Ecoato	Agridulce
MU18 ^a	Murcia	Ecoato	Agridulce
MU19 ^a	Murcia	Ecoato	Agridulce
MU20 ^a	Murcia	Ecoato	Agridulce
MU21 ^a	Murcia	Ecoato	Agridulce
MU22 ^d	Murcia	J. Antonio Navarro	Dulce
MU23 ^d	Murcia	J. Antonio Navarro	Dulce
MU24 ^d	Murcia	J. Antonio Navarro	Dulce
MU25 ^d	Murcia	La Odalisa	Dulce
MU26 ^d	Murcia	La Odalisa	Dulce
MU27 ^d	Murcia	La Odalisa	Dulce
MU28 ^d	Murcia	Alteza	Dulce
MU29 ^d	Murcia	Alteza	Dulce
MU30 ^d	Murcia	Alteza	Dulce
MU31 ^d	Murcia	Alteza	Dulce
MU32 ^d	Murcia	Alteza	Dulce
MU33 ^d	Murcia	Alteza	Dulce
MU34 ^d	Murcia	FG	Dulce
MU35 ^d	Murcia	FG	Dulce
MU36 ^d	Murcia	FG	Dulce
MU37 ^d	Murcia	FG	Dulce
MU38 ^d	Murcia	FG	Dulce
MU39 ^d	Murcia	FG	Dulce
MU40 ^d	Murcia	Pichi	Dulce
MU41 ^d	Murcia	Pichi	Dulce
MU42 ^d	Murcia	Pichi	Dulce

Tabla III.1. Continuación.

Código	Origen	Marca	Tipo
MU43 ^P	Murcia	J. Antonio Navarro	Picante
MU44 ^P	Murcia	J. Antonio Navarro	Picante
MU45 ^P	Murcia	J. Antonio Navarro	Picante
MU46 ^P	Murcia	La Odalisa	Picante
MU47 ^P	Murcia	La Odalisa	Picante
MU48 ^P	Murcia	La Odalisa	Picante
MU49 ^P	Murcia	Exportación Cervantes	Picante
MU50 ^P	Murcia	Exportación Cervantes	Picante
MU51 ^P	Murcia	Exportación Cervantes	Picante
MU52 ^P	Murcia	Exportación Cervantes	Picante
MU53 ^P	Murcia	Exportación Cervantes	Picante
MU54 ^P	Murcia	Exportación Cervantes	Picante
MU55 ^P	Murcia	Alteza	Picante
MU56 ^P	Murcia	Alteza	Picante
MU57 ^P	Murcia	Alteza	Picante
MU58 ^P	Murcia	Alteza	Picante
MU59 ^P	Murcia	Alteza	Picante
MU60 ^P	Murcia	Alteza	Picante
MU61 ^P	Murcia	Coato	Picante
MU62 ^P	Murcia	Coato	Picante
MU63 ^P	Murcia	Coato	Picante
MU64 ^P	Murcia	Choricero oro	Picante
MU65 ^P	Murcia	Choricero oro	Picante
MU66 ^P	Murcia	Choricero oro	Picante
MU67 ^P	Murcia	Choricero oro	Picante
MU68 ^P	Murcia	Choricero oro	Picante
MU69 ^P	Murcia	Choricero oro	Picante
MU70 ^P	Murcia	Pichi	Picante
MU71 ^P	Murcia	Pichi	Picante
MU72 ^P	Murcia	Pichi	Picante

En las tablas III.2, III.3 y III.4 se incluyen las muestras que se han estudiado considerando distintos tipos de variables. La tabla III.2 presenta las muestras que se han estudiado considerando el perfil metálico. En la III.3 se han incluido las muestras estudiadas por espectrofotometría UV-Vis, determinando el parámetro ASTA, y las coordenadas CIELab. La tabla III.4 incluye las muestras en las que se ha calculado el parámetro SHU.

Tabla III.2. Muestras analizadas mediante ICP-AES.

Origen	Tipo	n° de muestras	Códigos
Extremadura	Agridulce	24	EX01 ^a -EX24 ^a
Extremadura	Dulce	24	EX25 ^d -EX48 ^d
Extremadura	Picante	24	EX49 ^p -EX72 ^p
Murcia	Agridulce	21	MU01 ^a -MU21 ^a
Murcia	Dulce	21	MU22 ^d -MU42 ^d
Murcia	Picante	30	MU43 ^p -MU72 ^p

Tabla III.3. Muestras analizadas mediante espectroscopia UV-Vis.

Origen	Código	Tipo
Extremadura	EX01 ^a - EX13 ^a ; EX15 ^a -EX18 ^a ; EX23 ^a -EX24 ^a	Agridulce
Extremadura	EX25 ^d - EX30 ^d ; EX33 ^d - EX34 ^d ; EX37 ^d - EX48 ^d ;	Dulce
Extremadura	EX49 ^p - EX72 ^p	Picante
Murcia	MU03 ^a - MU07 ^a ; MU9 ^a -MU11 ^a ; MU15 ^a -MU20 ^a	Agridulce
Murcia	MU22 ^d -MU25 ^d ; MU27 ^d -MU31 ^d ; MU35 ^d -MU38 ^d ; MU42 ^d	Dulce
Murcia	MU45 ^p - MU48 ^p ; MU50 ^p ; MU53 ^p - MU55 ^p ; MU66 ^p ; MU57 ^p -MU71 ^p	Picante

Tabla III.4. Muestras analizadas para determinar el parámetro SHU.

Origen	Código	Tipo
Extremadura	EX49 ^p - EX72 ^p	Picante
Murcia	MU45 ^p - MU48 ^p ; MU50 ^p ; MU53 ^p - MU55 ^p ; MU66 ^p ; MU57 ^p -MU710 ^p	Picante

III.2. Procedimientos

III.2.1. Determinación de la humedad en muestras de pimentón

La humedad de las muestras se ha determinado siguiendo las recomendaciones de la AOAC para muestras vegetales [4]. En primer lugar se colocó la capsula destapada y la tapa durante 1 hora a 105°C, y empleando unas pinzas se trasladó a un desecador, donde se dejó enfriar durante 30 minutos; a continuación se pesó con una precisión de 0,1 mg. A continuación, se pesan 2 g de muestra, se colocan en la capsula destapada y se introducen en la estufa a 105°C durante 4 horas. Una vez transcurrido el tiempo se tapa la capsula y se lleva con unas pinzas al desecador, donde se deja durante 30 minutos para asegurar que se enfría hasta temperatura ambiente. Se pesa la muestra con la capsula tapada dos veces con un intervalo de 2 minutos para verificar que no existen variaciones en la pesada. Puesto que la medida de humedad de la muestra se basa en la determinación gravimétrica de la pérdida de masa correspondiente al agua presente, la forma de calcularla es la siguiente:

$$\% \text{Humedad} = \frac{m_2 - m_3}{m_2 - m_1} \cdot 100 \quad (3.1)$$

donde m_1 es la masa en gramos de la capsula vacía y de su tapa; m_2 es la masa en gramos de la capsula y la tapa con la muestra tapada antes del secado, y m_3 es la masa en gramos de la cápsula tapada con la muestra desecada.

III.2.2. Determinación de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn mediante ICP-AES

Para la determinación de los elementos anteriormente citados es preciso llevar a cabo la mineralización de las muestras. Se aplicaron tres métodos de digestión distintos por vía húmeda utilizando distintas mezclas de ácidos. Una descripción de cada uno de ellos se detalla a continuación.

▪ *Método 1 (M1)*

En este procedimiento se utiliza una mezcla $\text{HNO}_3/\text{H}_2\text{SO}_4$ (15:2) para llevar a cabo la digestión. Se coloca 1g de pimentón en un vaso de precipitado de 100 ml y se añaden 2 ml de H_2SO_4 , 5 ml HNO_3 y se calienta a 90-100°C en un baño de arena. El resto de HNO_3 se va añadiendo paulatinamente. Se calienta hasta una reducción de aproximadamente el 95% del volumen inicial y hasta que deje de observarse desprendimiento de vapores rojos de óxido nitroso.

▪ *Método 2 (M2)*

En este caso se utiliza una mezcla $\text{HNO}_3/\text{H}_2\text{O}_2$ (4:1) para llevar a cabo la digestión. Se coloca 1g de pimentón en un vaso de 100 ml y se añaden 5 ml de HNO_3 y 5 ml de H_2O_2 , se calienta a 90-100°C en un baño de arena. Se añaden 15 ml más de HNO_3 , paulatinamente, y se sigue calentando hasta que dejen de observarse vapores rojos de óxido nitroso.

▪ *Método 3 (M3)*

Para la mineralización de la muestra en este caso se recurre a una mezcla de $\text{HNO}_3/\text{HClO}_4$ (4:1). Se coloca 1 g de pimentón en un vaso de precipitado de 100 ml junto a 5 ml de HClO_4 y 5 ml de HNO_3 gota a gota para evitar reacciones bruscas. Se calienta durante 10 minutos a unos 90-100°C y a continuación se añaden 15 ml de HNO_3 , paulatinamente. Se calienta hasta que dejan de observarse vapores rojos de óxido nitroso.

Todas las disoluciones resultantes de la aplicación de estos procedimientos de mineralización se dejan enfriar a temperatura ambiente, se filtran (0,45 μm) y se transfieren a un matraz aforado de 50 ml, enrasando con agua Milli-Q.

Las condiciones instrumentales para la medida mediante ICP-AES se presentan en la tabla III.5.

Tabla III.5. Condiciones instrumentales ICP-AES

RF frecuencia	40,68 MHz
Potencia operacional	1000 W
Observación	Radial
Flujo de argón (plasma)	12 L min ⁻¹
Tipo de nebulizador	Meinhard
Flujo aspiración muestra	1 mL min ⁻¹
Autosampler	AS500
Elemento	λ / nm
Al	167,02
B	208,96
Ca	317,93
Cu	324,75
Fe	259,94
K	766,49
Mg	279,55
Mn	257,61
Na	589,59
Ni	221,65
P	178,23
Pb	220,35
Sr	346,45
Zn	213,86

III.2.3. Medidas espectrofotométricas

Los parámetros ASTA y las coordenadas CIELab se han calculado a partir de medidas de absorbancia en el intervalo UV-Vis, realizadas con cubetas de cuarzo de 1,0 cm de paso de luz. A continuación se detallan los procedimientos utilizados en cada caso.

III.2.3.1. Determinación del parámetro ASTA

Para determinar el color en muestras de pimentón se sigue el procedimiento ASTA 20.1 [4]. Se pesan entre 0,07 y 0,1 g de pimentón, se disuelven en acetona, se agita, se deja reposar, y se enrasa con el disolvente a 100 mL. Antes de medir la absorbancia a 460 nm, se vuelve a agitar y se dejan depositar las partículas 2 minutos.

Es necesario realizar un cálculo de corrección del equipo, I_f , tal y como se muestra en la ecuación 3.2. Para ello se mide la absorbancia del patrón consistente en una disolución de sulfato de amonio y cobalto (II) hexahidratado ($34,96 \text{ g L}^{-1}$) y dicromato de potasio ($0,3005 \text{ g L}^{-1}$), en disolución de ácido sulfúrico 1,8 M. El cálculo del factor de corrección se realiza según la siguiente ecuación:

$$I_f = \frac{A_p}{A} \quad (3.2)$$

siendo A_p la absorbancia teórica del patrón a 460 nm (0,6) y A la absorbancia experimental del patrón a 460 nm.

El cálculo de grados ASTA se realiza según la ecuación 3.3:

$$\text{Color ASTA} = \frac{A_m \times 16.4 \times I_f}{m} \quad (3.3)$$

donde A_m es la absorbancia de la muestra a 460 nm y m la masa de la muestra en gramos.

III.2.3.2. Registro de espectros para el cálculo de las coordenadas CIELab.

Para calcular las coordenadas CIELab se registran los espectros UV-Vis de las muestras en el intervalo de longitudes de onda 360 – 800 nm, con un intervalo de 1 nm. Las muestras se preparan de manera similar a la indicada en el apartado III.2.3.1.

III.2.3.3. Registro de espectros para el cálculo del parámetro SHU.

Se pesa 1 g de muestra de pimentón y se añade un volumen de acetonitrilo de 20 mL, tratándose durante 20 minutos en un baño de ultrasonido a 50°C. Se toman 0,5

mL y se llevan a un volumen final de 10 mL con acetonitrilo. Los espectros se registran en el intervalo de longitudes de onda de 215 a 800 nm. Como las bandas características de los capsaicinoides se encuentran en el intervalo de 230 a 280 nm [6], los cálculos se han realizado utilizando los valores de absorbancia comprendidos entre 215 y 315 nm [6]. Como la mayoría de los carotenoides absorben a longitudes de onda comprendidas entre 560-580 nm, este intervalo de longitudes de onda no es útil [7,8].

III.2.4. Determinación de capsaicinoides mediante HPLC-UV

En primer lugar se realiza una extracción de los capsaicinoides. Se pesa 1 g de muestra de pimentón y se añade un volumen de 20 mL de acetonitrilo. Se aplica ultrasonido durante 20 minutos a 50°C. A continuación, se extraen los capsaicinoides de la disolución obtenida mediante SPE con cartuchos de 200 mg de C18 [8]. Para acondicionar el cartucho se hacen pasar 5 mL de acetonitrilo y después 5 mL de agua. 1 mL de la disolución inicial obtenida se diluye con agua hasta 10 mL. Estos 10 mL se pasan a través del cartucho de C18 acondicionado. Los capsaicinoides se eluyen con 5 mL de acetonitrilo. La disolución se microfiltra a través filtros de nylon de 0,45 µm y se inyecta en el cromatógrafo líquido. Las condiciones cromatográficas se presentan en la tabla III.6.

Tabla III.6. Condiciones cromatográficas para la determinación de capsaicinoides en muestras de pimentón.

Fase móvil	Acetonitrilo/agua (63:37)
Fase estacionaria	C18
Flujo de la fase móvil	1,1 mL min ⁻¹
Volumen de inyección	20 µL
Temperatura de la columna	43°C
Longitud de onda de detección	280 nm

III.3. Bibliografía

1. Standard conditions, *Analytical Methods for Atomic Absorption Spectrophotometry*, PERKIN ELMER, **1982**, Norwalk, Connecticut, U.S.A.
2. Reglamento (CE) núm. 982/2007 21-08-2007. (DOCE 22-08-2007, L. 217, P. 0022-0023).
3. Reglamento (CE) núm. 464/2001 07-03-2001. (DOCE 8-03-2001, L. 66, P. 0029-0030).
4. Official Methods of Analysis. A.O.A.C., 19th, 2012.
5. D.L.B. Wetzel. *Instrumental Methods in Food and Beverage Analysis*. Elsevier, Amsterdam, 1998.
6. U. Schweiggert, R. Carle, A. Schieber. Characterization of major and minor capsaicinoids and related compounds in chili pods (*Capsicum frutescens* L.) by high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Analytica Chimica Acta*, **2006**, 557, 236-244.
7. Y. M. A. Naguib. Antioxidant activities of astaxanthin and related carotenoids. *Journal of Agricultural and Food Chemistry*, **2000**, 48, 1150-1154.
8. B. Röhl. Carotenoid and retinoid – two pigments in a gecko eye lens. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **2000**, 125, 105-112.

IV. Discusión y resultados.

IV.1. Determinación de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn por ICP-AES

Se han propuesto tres métodos de mineralización para la determinación de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn en muestras de pimentón, llevándose a cabo una comparación de los mismos, con el propósito de seleccionar el más adecuado. Dichos métodos, descritos en la parte experimental III.2.1.1, consisten en una digestión por vía húmeda utilizando las mezclas ácidas: H₂SO₄:HNO₃ (M1), H₂O₂:HNO₃ (M2) y HClO₄:HNO₃ (M3).

IV.1.1. Comparación de los métodos de mineralización

La comparación entre métodos de mineralización, se llevó a cabo mediante un estudio de la exactitud, ya que este parámetro tiene en cuenta las incertidumbres relacionadas con el sesgo inherente al método [1]. El protocolo LGC/VAM [2] propone varios métodos para determinar la incertidumbre relacionada con la veracidad: el análisis de un material de referencia, la comparación con un método de referencia y los ensayos de recuperación con muestras fortificadas.

En el caso del pimentón no existe ningún tipo de material de referencia certificado, con lo que la comparación se realiza mediante el análisis de las recuperaciones obtenidas tras aplicar cada método de digestión a muestras fortificadas [3]. Para ello, se fortificaron distintas porciones de una muestra de pimentón con los elementos a determinar, Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn. Se realizaron adiciones a tres niveles de concentración, con el fin de cubrir los rangos esperados a partir de una serie de análisis previos en distintas muestras de pimentones. Se consideraron tres niveles de adición, 1, 2, y 3, correspondientes al 80, 100 y 120 % de los valores esperados para cada elemento, respectivamente. El análisis se realizó por triplicado. Los resultados de las recuperaciones por niveles y métodos se recogen en la tablaIV.1.

Tabla IV.1. Recuperaciones en muestras fortificadas a tres niveles

Elemento	Nivel de adición	Recuperación (%)		
		M1	M2	M3
Al	1	100,72	99,79	97,84
	2	96,91	106,31	99,41
	3	99,25	100,03	100,01
B	1	94,03	105,45	97,02
	2	100,94	105,00	97,20
	3	100,29	101,67	95,61
Ca	1	99,68	112,10	103,54
	2	107,43	99,94	99,19
	3	99,46	98,00	96,20
Cu	1	91,62	89,73	96,34
	2	92,18	82,06	97,35
	3	87,87	95,49	97,53
Fe	1	86,88	101,08	102,67
	2	81,37	116,19	100,75
	3	77,46	99,61	99,59
K	1	101,00	92,41	99,44
	2	99,38	104,74	100,35
	3	99,58	106,13	100,23
Mg	1	113,55	100,12	101,63
	2	105,47	99,68	99,46
	3	106,00	99,30	102,16
Mn	1	109,73	81,94	98,21
	2	80,51	94,70	96,67
	3	88,92	95,45	97,74
Na	1	100,50	78,01	99,05
	2	93,78	80,21	96,56
	3	99,07	94,64	98,62
Ni	1	101,21	94,79	97,01
	2	95,36	97,43	95,89
	3	81,51	85,71	100,04
P	1	103,45	99,66	100,76
	2	88,47	101,48	99,51
	3	96,97	106,16	101,09
Pb	1	92,04	83,06	99,58
	2	92,93	104,07	102,13
	3	85,30	90,04	99,87
Sr	1	105,5	77,47	99,09
	2	86,33	87,01	96,48
	3	83,53	82,11	94,22
Zn	1	107,40	113,67	95,25
	2	104,66	102,79	98,30
	3	117,13	125,93	94,12

M1: H₂SO₄-HNO₃; M2: H₂O₂-HNO₃; M3: HClO₄-HNO₃

La comparación de los resultados obtenidos por los tres métodos se ha realizado mediante un ANOVA de dos factores anidados. El método de digestión (factor A) varía en $i = (1 \dots a)$. Por cada i método de digestión se tienen $j = (1 \dots b)$ niveles de fortificación (factor B) y se realizan $k = (1 \dots r)$ replicados. Esta serie de experimentos se puede esquematizar en la figura IV.1, donde R_{ijk} es la recuperación del k -ésimo replicado para el nivel j del método i , \bar{R}_{ij} es el valor medio de las recuperaciones R_{ijk} , \bar{R}_i la recuperación media del método i y $\bar{\bar{R}}$ es la media global de todas las recuperaciones.

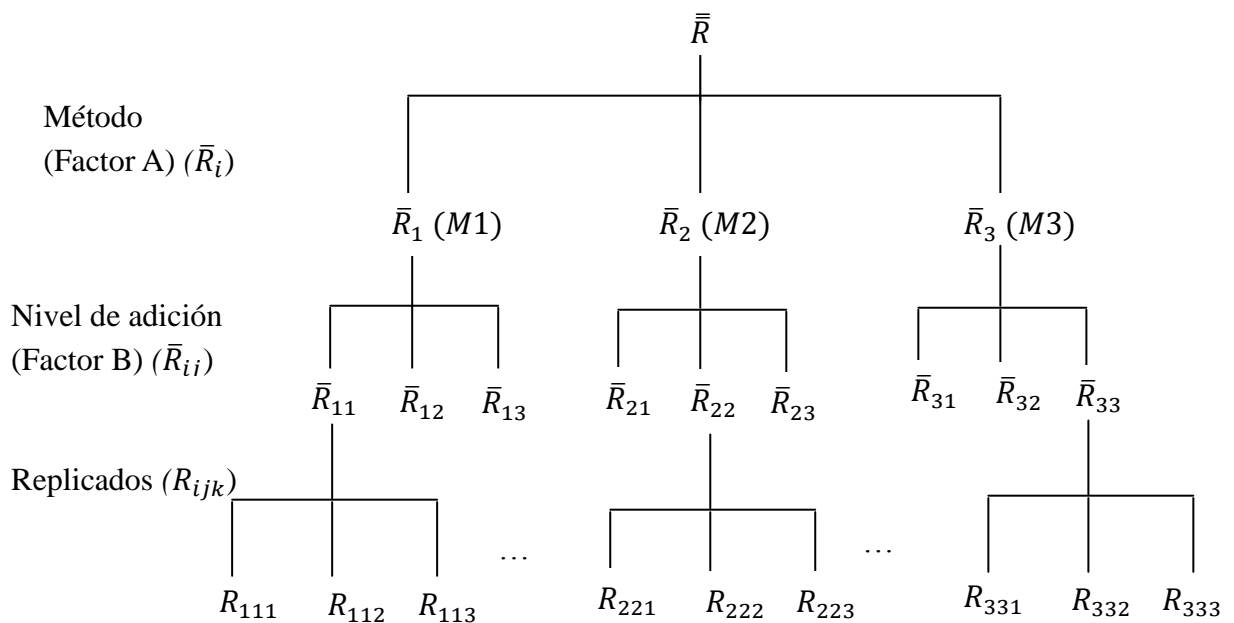


Figura IV.1. ANOVA de dos factores anidados para el estudio de la recuperación.

La recuperación R_{ijk} de cada elemento se calcula según la ecuación 4.1

$$\bar{R}_{ijk} = \frac{C_{observada} - C_{nativa}}{C_{añadida}} \times 100 \quad (4.1)$$

donde C_{nativa} es la concentración del analito en la muestra sin fortificar, $C_{observada}$ la determinada en la muestra fortificada y $C_{añadida}$ la concentración añadida de analito.

Para comparar los métodos de mineralización a partir de los cálculos de ANOVA y detectar sesgos significativos, se compara la varianza entre métodos (S^2_{metodo})

(ecuación 4.2) con la de error puro (S^2_{rep}) (ecuación 4.3), mediante una prueba F de Fisher (ecuación 4.4) de dos colas.

$$S^2_{\text{método}} = \frac{rb \sum_{i=1}^a \left(\bar{R}_i - \bar{R} \right)^2}{a-1} \quad (4.2)$$

$$S^2_{rep} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r \left(R_{ijk} - \bar{R}_{ij} \right)^2}{ab(r-1)} \quad (4.3)$$

El valor F experimental se calcula según la expresión:

$$F = \frac{S^2_{\text{método}}}{S^2_{rep}} \quad (4.4)$$

Si el valor calculado de F es mayor que el tabulado para $\alpha=0,05$ y con $a-1$ y $ab(r-1)$ grados de libertad (anexo A2.3), existe entonces un sesgo significativo debido a los métodos. Esto implica que al menos una de las recuperaciones medias, \bar{R}_i , difiere del resto. En la tabla IV.2. se presentan los resultados obtenidos en estas pruebas de significación. Como se observa, no hay diferencias entre los métodos para los elementos B, Ca, K Mn y Ni. En la tabla IV.2., el estadístico F indica la existencia de un sesgo significativo debido al método para los elementos Al, Cu, Fe, Mg, Na, P, Pb, Sr y Zn. Para detectar el método de digestión que produce el sesgo se realiza la prueba de la genuina diferencia significativa (HSD) de Tukey [4], comparando las recuperaciones medias obtenidas por cada método.

Como puede observarse que en el caso del Sr existen diferencias entre todos los métodos. Fe y Mg presentan diferencias entre M1 y los demás métodos. Al presenta las diferencias significativas entre M3 y el resto, y para Cu, Pb y Zn es el método M3 el que presenta diferencias significativas. Na y P sólo presentan diferencias entre M1 y M2. Estas diferencias deben ser analizadas con más detalle.

Tabla IV.2. Cálculos ANOVA y análisis *post-hoc*.

Elemento	Cálculos ANOVA			Análisis <i>post-hoc</i>		
	S^2_{metodo}	S^2_{rep}	F_{exp}	M1-M2	M1-M3	M3-M2
Al	27,4	5,8	4,72^a			
B	139,7	96	1,46			
Ca	42,7	59	0,72			
Cu	162,4	8,5	19,11^a			
Fe	1423	24,6	57,85^a			
K	3,6	9,4	0,38			
Mg	193,8	6,6	29,36^a			
Mn	108,9	74,4	1,46			
Na	558,4	5,8	96,28^a			
Ni	74,4	85,6	0,87			
P	88,3	21,7	4,07^a			
Pb	270,6	15,1	17,92^a			
Sr	489,5	16,9	28,96^a			
Zn	815,6	123,5	6,60^a			

^aDiferencias significativas para $F_{\text{exp}} > F_{\text{tab}}(0,05,2,18) = 3,5$

Las diferencias pueden aparecer en uno o varios niveles de concentración. Las recuperaciones deben tener valores del orden del 100% para todos los niveles de fortificación, dando un resultado global del 100%. Para poder detectar la presencia de niveles de fortificación con recuperaciones diferentes del 100% se representa la recuperación media de cada nivel *j* para cada método *i* con su desviación estándar.

Las figuras IV.2.- IV.15. incluyen los valores de recuperación por niveles y métodos para los distintos elementos

En el caso del Al, en la figura IV.2., se aprecian diferencias significativas en el M2 para el nivel 2 de fortificación. Para el caso del B, Mn, Ni y Ca no existen diferencias entre los métodos de digestión. Sin embargo, si se realiza un análisis *post-*

hoc, se puede ver como en el caso del Mn sí existen diferencias significativas en los valores de recuperación para el nivel 1 de fortificación del 80% del M2 y en el nivel 2 del M1. Para el Ni y el Ca, el análisis *post-hoc* revela que no existen diferencias significativas entre niveles. Como se observa en la figura IV.4. los valores de recuperación del nivel 1 del Ca para el M2 difieren del resto. En el caso del Ni (figura IV.11.) los valores de recuperación para el nivel 3, correspondiente al 120% de fortificación, difieren del resto. En el caso del Cu, Pb y Zn, de acuerdo con la tabla IV.2., vemos que existen diferencias del M3 con respecto al 1 y al 2, lo que se confirma en las figuras IV.5., IV.13., y IV.15. En el caso del Fe, en la figura IV.6. se observa cómo el M1 presenta diferencias con respecto al 2 y al 3. Con el M1 se obtienen valores de recuperaciones más bajos que para los M2 y M3. Considerando los valores del Fe para el M2 se aprecia que son superiores a 100% para el nivel intermedio de fortificación. El análisis *post-hoc* para el K no muestra diferencias entre los métodos, pero al considerar los resultados para cada nivel, se aprecia que en el caso de los niveles 1 y 3 del M2 hay diferencias importantes, tal y como se aprecia en la figura IV.7. En el caso del Mg, el análisis *post-hoc* da resultados similares a los obtenidos con el Fe, en el que el M1 da resultados diferentes a los otros dos; no obstante, en este caso, los valores del M1 son mayores, lo que se puede ver en la figura IV.8. Los resultados del análisis *post-hoc* para el Na muestran diferencias entre los M1 y M2. En la figura IV.10. se comprueba que los niveles 1 y 2 del M2 presentan diferencias con los demás, ya que son muy bajos. En el caso del P, existen diferencias entre M1 y M2, y en la figura IV.12. se aprecia cómo el nivel 2 del M2 da valores diferentes. En el caso del Sr, los tres métodos de digestión presentan diferencias entre sí, apreciándose en la figura IV.14. que el nivel 1 del M2 da valores especialmente bajos. Análogamente, los niveles 2 y 3 de M2 y M1 son bajos.

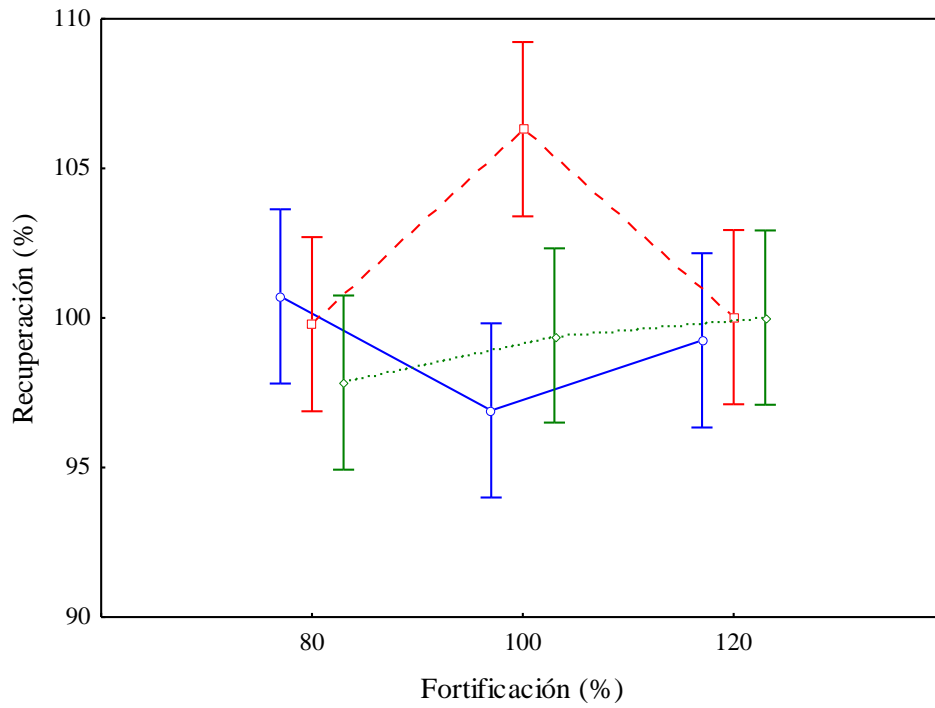


Figura IV.2. Recuperación para Al en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···)

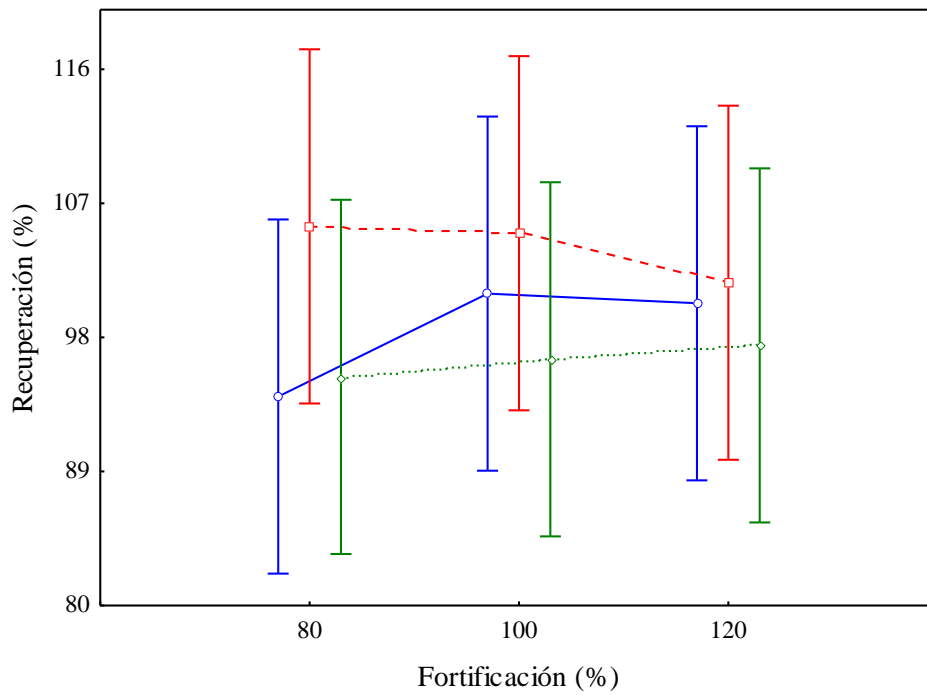


Figura IV.3. Recuperación para B en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

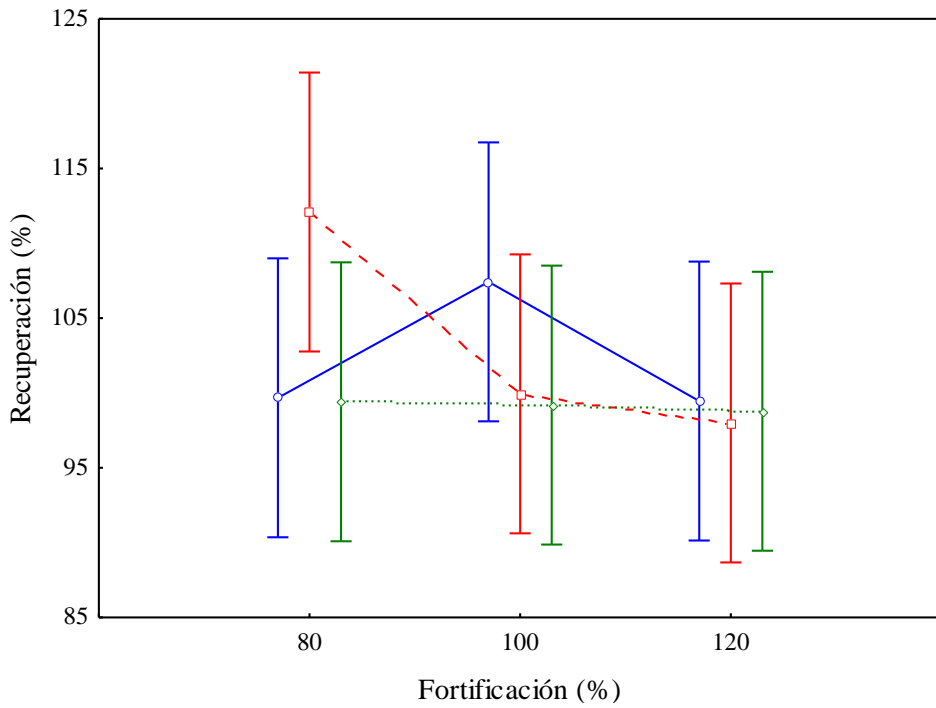


Figura IV.4. Recuperación para Ca en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

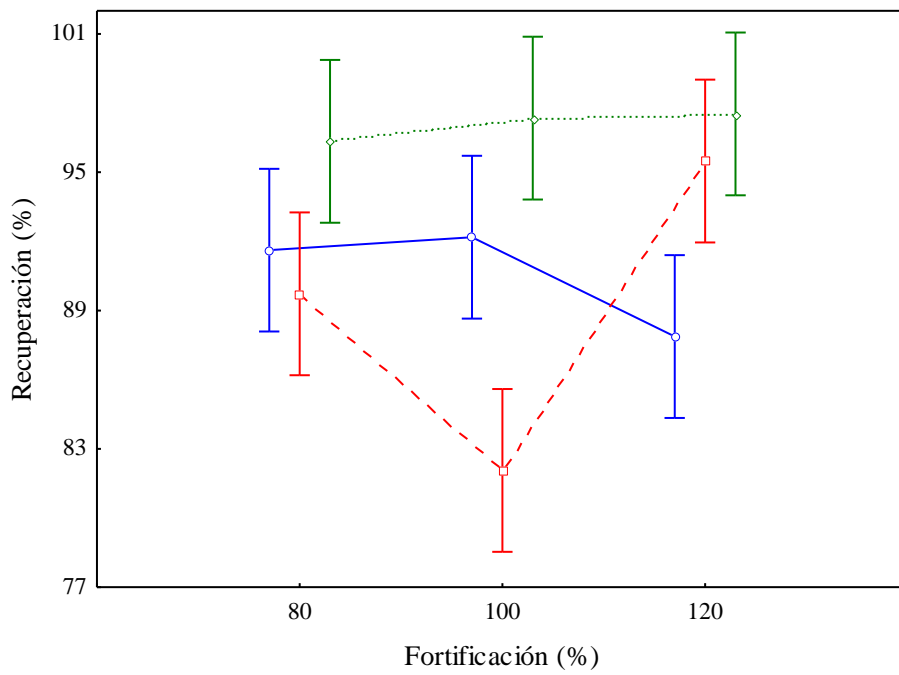


Figura IV.5. Recuperación para Cu en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

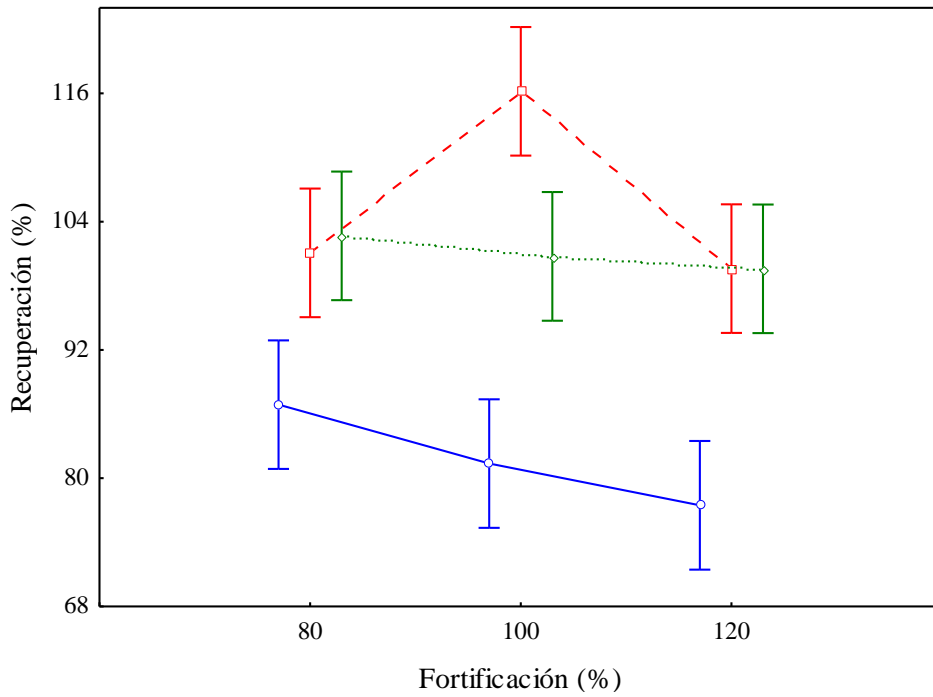


Figura IV.6. Recuperación para Fe en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

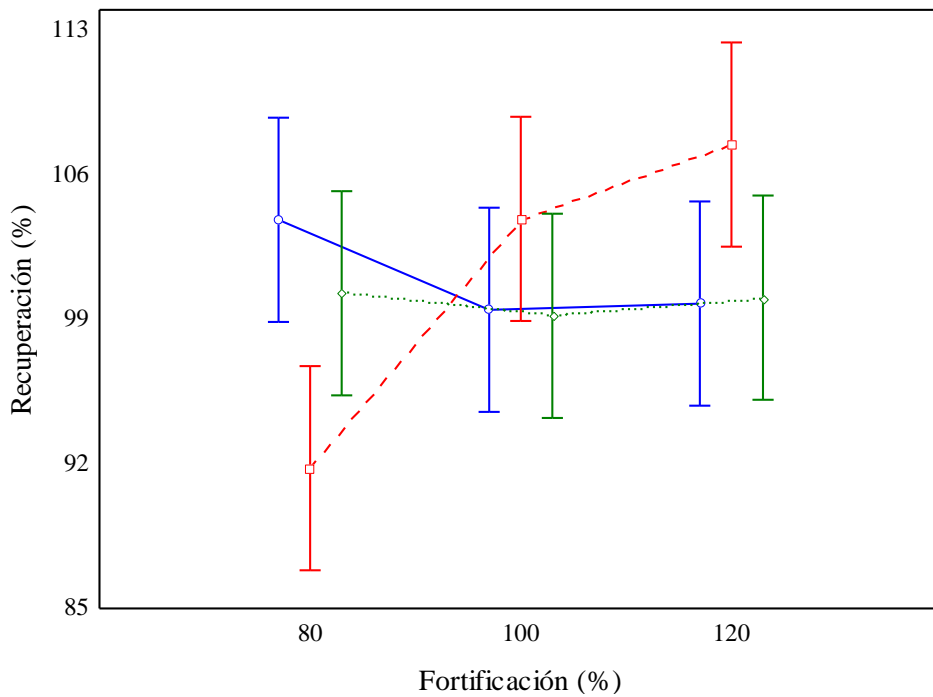


Figura IV.7. Recuperación para K en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

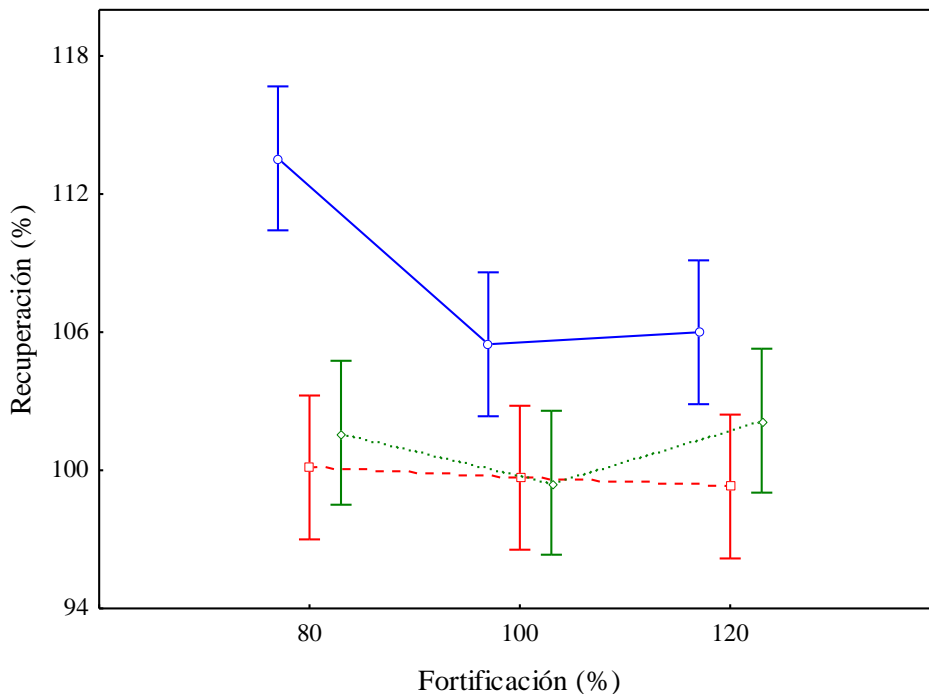


Figura IV.8. Recuperación para Mg en los distintos niveles de fortificación. M1 (—), M2 (- - -), M3 (· · ·).

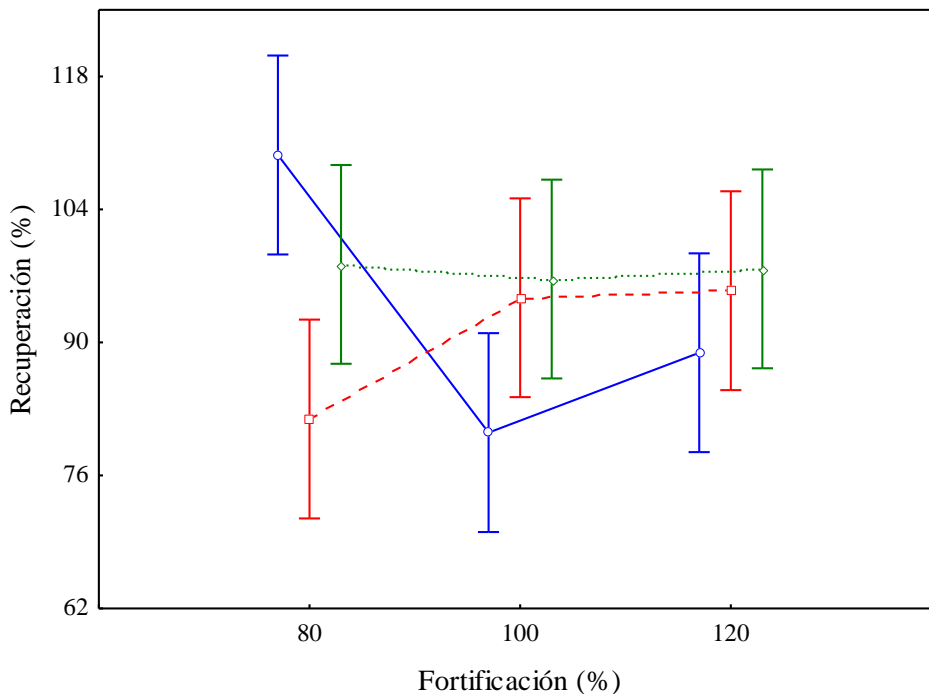


Figura IV.9. Recuperación para Mn en los distintos niveles de fortificación. M1 (—), M2 (- - -), M3 (· · ·).

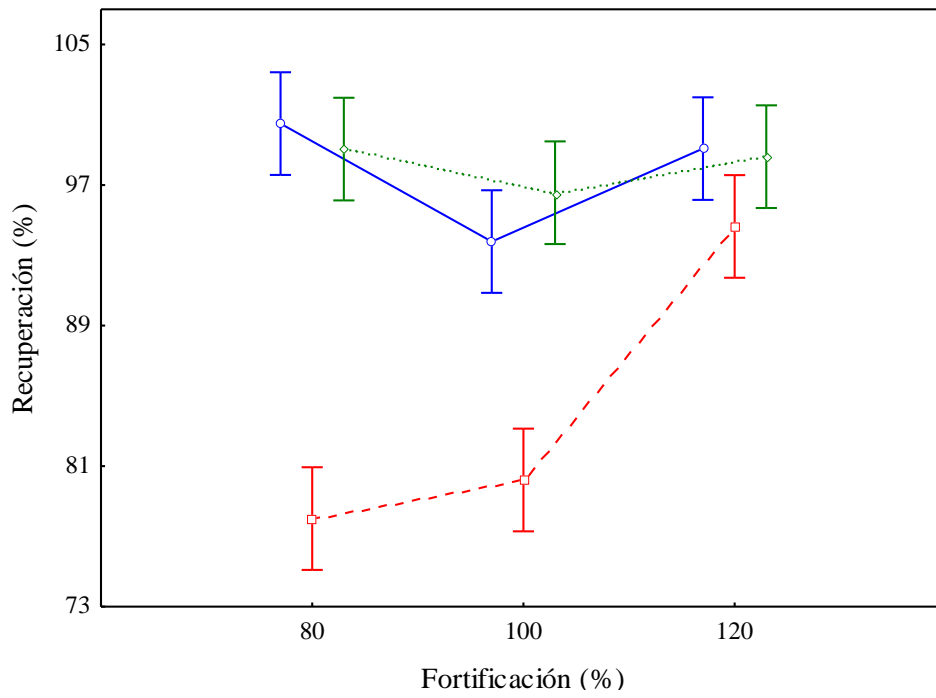


Figura IV.10. Recuperación para Na en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

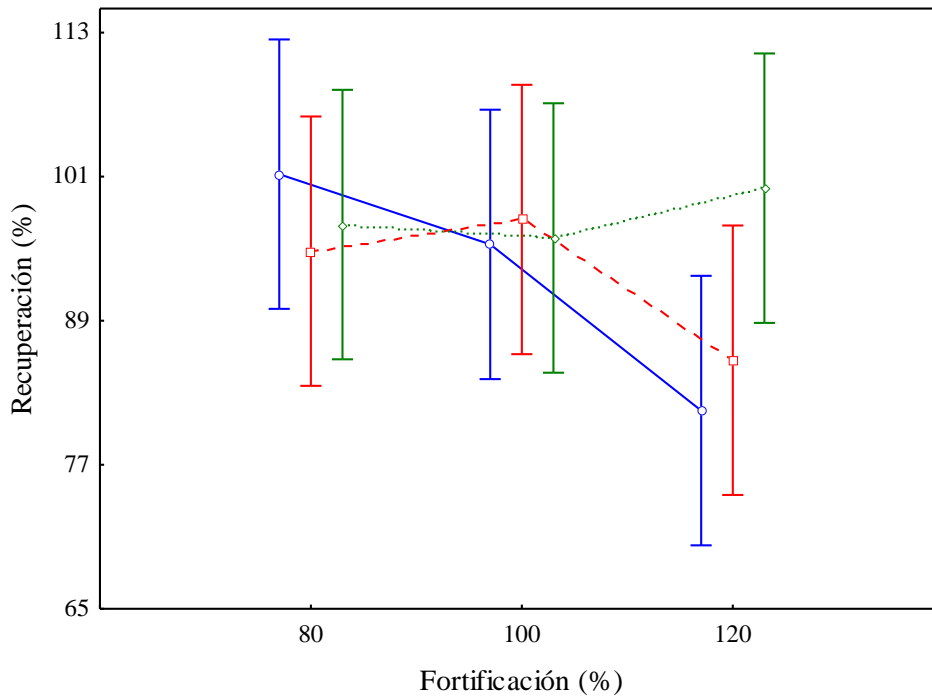


Figura IV.11. Recuperación para Ni en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

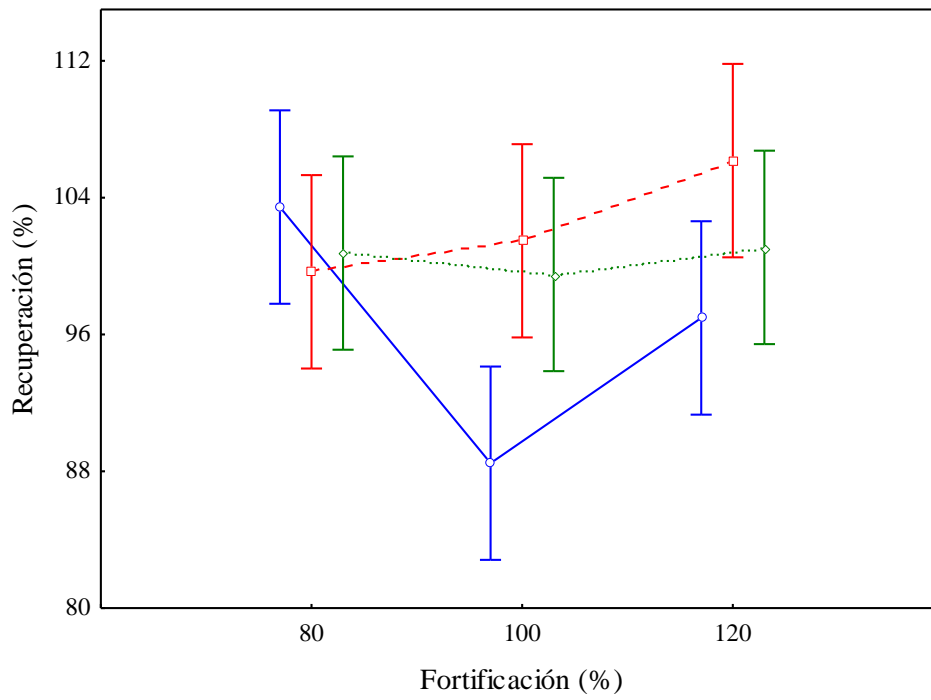


Figura IV.12. Recuperación para P en los distintos niveles de fortificación. M1 (—), M2 (---), M3 (···).

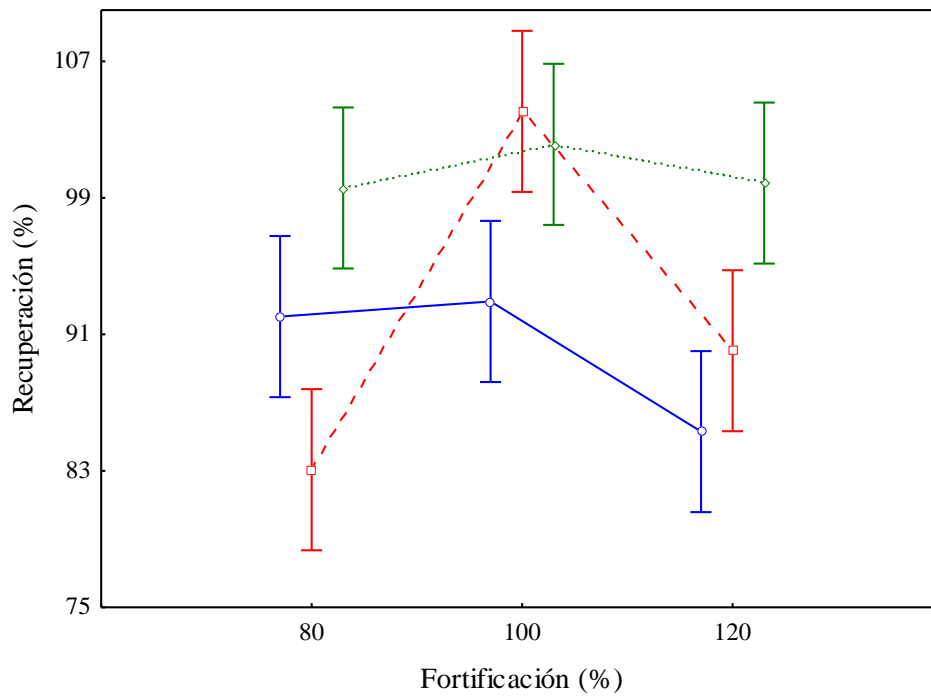


Figura IV.13. Recuperación para Pb en los distintos niveles de fortificación. M1 (—), M2 (---), M3 (···).

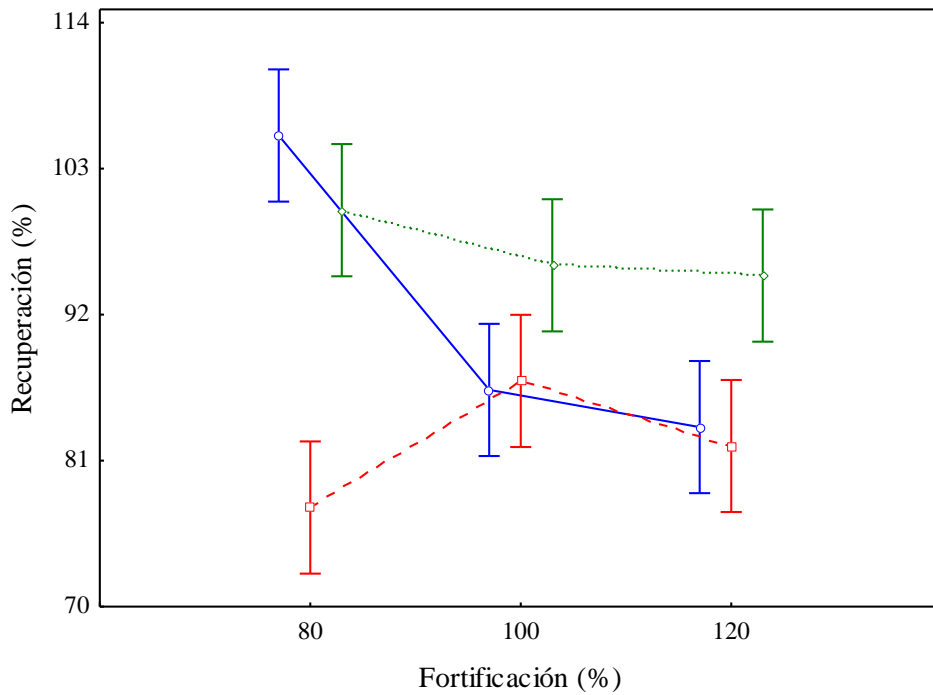


Figura IV.14. Recuperación para Sr en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

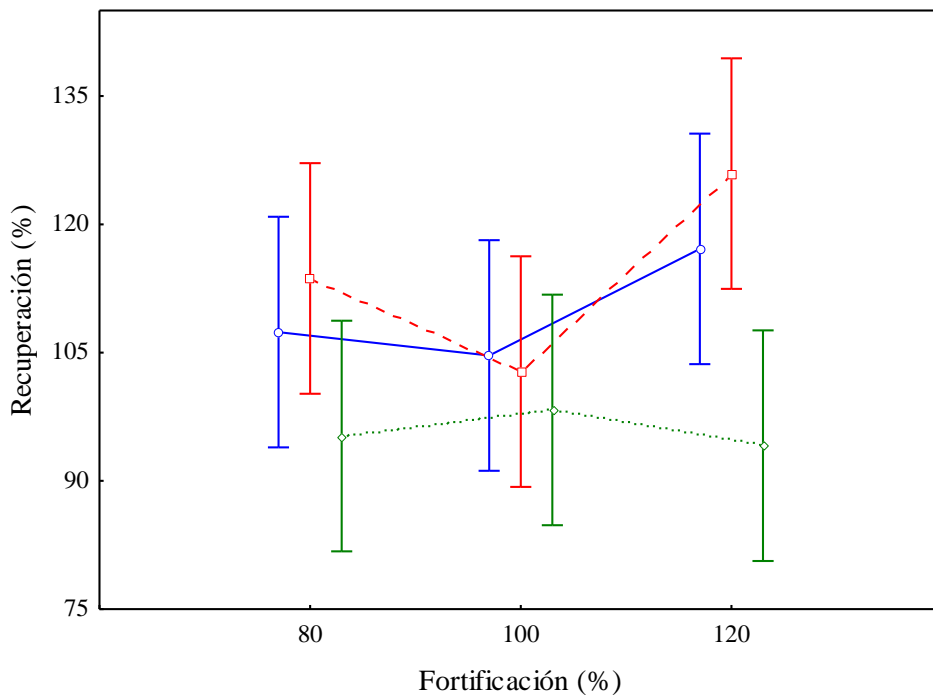


Figura IV.15. Recuperación para Zn en los distintos niveles de fortificación.
M1 (—), M2 (---), M3 (···).

En el estudio sobre las recuperaciones es importante conocer la incertidumbre asociada a cada uno de los valores obtenidos para los distintos niveles de recuperación en cada método.

Cada valor de \bar{R}_{ij} tendrá asociada una incertidumbre $u(\bar{R}_{ij})$, que viene dada por la desviación estándar de las medidas de cada nivel ij . La incertidumbre correspondiente a los valores de \bar{R}_i tiene dos términos. Uno es la incertidumbre combinada $u_c(\bar{R}_i)$, que corresponde a la propagación de las incertidumbres de cada nivel ij (4.5).

$$u_c^2(\bar{R}_i) = \frac{1}{b^2} \left[\sum_{j=1}^b u^2(\bar{R}_{ij}) \right] \quad (4.5)$$

donde b es el número de niveles.

El otro término es la desviación estándar de la media \bar{R}_i , que es una incertidumbre de repetición pura, $u_{rep}(\bar{R}_i)$ (4.6).

$$u_{rep}^2(\bar{R}_i) = \frac{\sum_{j=1}^b (\bar{R}_{ij} - \bar{R}_i)^2}{b(b-1)} \quad (4.6)$$

La incertidumbre total para cada método, $u(\bar{R}_i)$, se calcula según:

$$u(\bar{R}_i) = \sqrt{u_c^2(\bar{R}_i) + u_{rep}^2(\bar{R}_i)} \quad (4.7)$$

La incertidumbre expandida se calcula multiplicando $u(\bar{R}_i)$ por un factor de cobertura, k , que para un 95% de nivel de confianza es igual a 2 [5].

$$U(\bar{R}_i) = 2 \cdot u(\bar{R}_i) \quad (4.8)$$

Los métodos se considerarán exactos cuando sus recuperaciones sean estadísticamente iguales a 100, es decir, cuando la siguiente desigualdad sea cierta.

$$|\bar{R}_i - 100| \leq U(\bar{R}_i) \quad (4.9)$$

Los resultados obtenidos se recogen en la tabla IV.3.

Como puede observarse para la mayoría de los valores obtenidos, el intervalo de confianza incluye el 100%. En el caso del Cu, Fe, Mn, Na, Ni, Pb, Sr y Zn, se obtienen valores de recuperación más próximos al 100% cuando se emplea el M3. Además, este método presenta menores valores de incertidumbre. De todos modos, a la hora de comparar métodos hay que considerar también otros criterios además de los puramente estadísticos, como los relacionados con la adecuación práctica del método [6].

Por un lado, la presencia de ácidos en la disolución final puede producir cierto efecto matriz, debido a diferencias de viscosidad entre las disoluciones de patrones y de las muestras. Por tanto, es útil eliminar la mayor cantidad posible de exceso de ácido y es mucho más fácil cuando se utiliza ácido perclórico o peróxido de hidrógeno en lugar de ácido sulfúrico ya que el punto de ebullición de este último es mayor. Además, el tratamiento nítrico-sulfúrico puede dejar residuos insolubles en forma de sulfatos de metales alcalinotérreos o de Pb. Por otra parte, el uso de peróxido de hidrógeno hace que la digestión no sea tan violenta como con ácido perclórico, pero la mezcla no es tan eficaz en el caso de este tipo de muestra pues presenta un alto contenido en materia grasa [7].

Tabla IV.3. Recuperaciones e incertidumbres de los distintos métodos de mineralización.

Elemento	Recuperación (%)		
	$\bar{R}_{M1} \pm U(\bar{R}_{M1})$	$\bar{R}_{M2} \pm U(\bar{R}_{M2})$	$\bar{R}_{M3} \pm U(\bar{R}_{M3})$
Al	99 ± 4	102 ± 5	99 ± 1
B	98 ± 13	104 ± 15	97 ± 3
Ca	102 ± 15	103 ± 11	99 ± 2
Cu	91 ± 6	89 ± 8	97 ± 3
Fe	82 ± 6	106 ± 14	101 ± 3
K	100 ± 4	101 ± 10	100 ± 2
Mg	108 ± 6	100 ± 5	101 ± 2
Mn	93 ± 23	91 ± 11	98 ± 4
Na	98 ± 5	84 ± 11	98 ± 3
Ni	93 ± 17	93 ± 15	98 ± 4
P	96 ± 10	102 ± 8	100 ± 1
Pb	90 ± 8	92 ± 13	101 ± 3
Sr	92 ± 14	82 ± 9	97 ± 3
Zn	110 ± 13	114 ± 24	96 ± 4

Por lo tanto, desde el punto de vista estadístico y práctico, el método más adecuado para la digestión y análisis de muestras de pimentón es el M3, siendo por eso el empleado en la mineralización de las muestras para la determinación de los catorce elementos en estudio.

IV.1.2. Efecto Matriz

Para estudiar el efecto matriz se realizó con fines comparativos el calibrado externo y el calibrado de adición patrón. En ausencia de efecto matriz, las pendientes en ambos casos deben ser iguales, hecho que se puede comprobar estadísticamente mediante un test t de Student. En primer lugar realizamos el cociente entre ambas pendientes y calculamos la desviación estándar asociada al mismo.

$$R = \frac{b_{CE}}{b_{AP}} \quad (4.10)$$

$$S_R = R \sqrt{\left(\frac{S_{b_{CE}}^2}{b_{CE}^2} + \frac{S_{b_{AP}}^2}{b_{AP}^2} \right)} \quad (4.11)$$

donde b_{CE} y b_{AP} son las pendientes de las rectas de calibrado externo y adición patrón, respectivamente, y $s_{b_{CE}}$ y $s_{b_{AP}}$ son sus desviaciones estándar correspondientes. Si ambas pendientes son estadísticamente iguales, su cociente, R , no debe ser significativamente distinto de 1. Calculando la t de Student como:

$$t_{cal} = \frac{|R - 1|}{S_R} \quad (4.12)$$

El valor calculado se compara con el obtenido para los grados de libertad efectivos de S_R y un 95% de nivel de confianza (anexo A2.1). Los grados de libertad efectivos se calculan a partir de la ecuación de Welch-Satterthwaite [8] como:

$$V_{eff} = \frac{S_R^4}{\frac{s_{b_{CE}}^4}{b_{AP}^4 (n_{CE} - 2)} + \frac{b_{CE}^4 s_{b_{AP}}^4}{b_{AP}^8 (n_{AP} - 2)}} \quad (4.13)$$

Los resultados se muestran en la tabla IV.4. En todos los casos el valor t_{cal} es menor que t_{tab} (v , 95%), con lo cual se concluye que no hay efecto matriz

Tabla IV.4. Estudio sobre la presencia de efecto matriz.

Elemento	b_{AP}	s_{bAP}	b_{CE}	s_{bCE}	t_{cal}	$t_{tab} (95\%)$
Al	974,4	23,4	1034,0	17,9	2,20	3,18
B	181,9	8,8	170,2	7,4	0,93	4,30
Ca	13,3	0,3	14,3	0,4	2,00	3,18
Cu	104,8	0,5	110,5	0,2	0,67	4,30
Fe	124,8	8,6	119,8	1,1	0,55	4,30
K	104,2	6,0	111,5	0,7	1,30	4,30
Mg	1192,3	18,3	1147,4	34,3	1,08	4,30
Mn	1364,7	22,7	1509,4	107,0	1,61	4,30
Na	122,1	1,8	110,6	3,1	2,74	3,18
Ni	1361,6	39,0	1247,9	59,1	1,38	3,18
P	27,9	0,25	27,4	0,23	1,58	2,78
Pb	389,5	13,8	348,0	11,4	1,97	3,18
Sr	120,7	4,6	109,4	6,3	1,23	3,18
Zn	168,7	7,4	178,7	7,7	1,01	3,18

b_{AP} : pendiente recta AP; s_{bAP} : desviación estándar recta AP; b_{CE} : pendiente recta CE; s_{bCE} : desviación estándar recta CE.

La representación gráfica de las rectas de calibrado externo junto a las de adición patrón para cada uno de los elementos analizados permite visualizar las posibles diferencias. Las figuras IV.16. – IV.29. muestran las rectas de adición patrón (AP) y calibrado externo (CE) para cada uno de los elementos determinados.

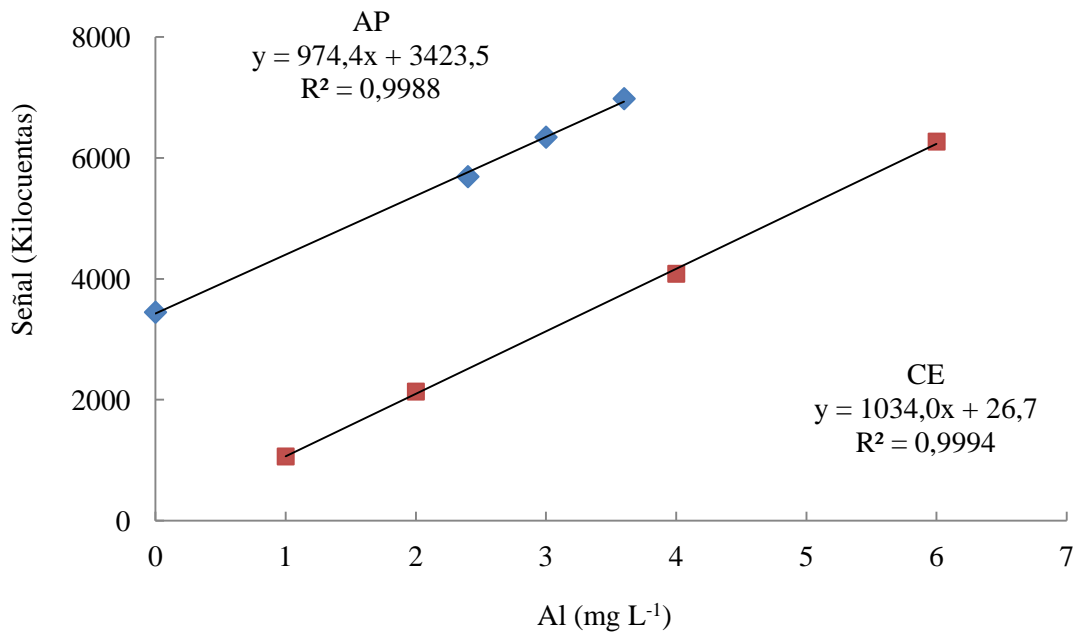


Figura IV.16. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de aluminio.

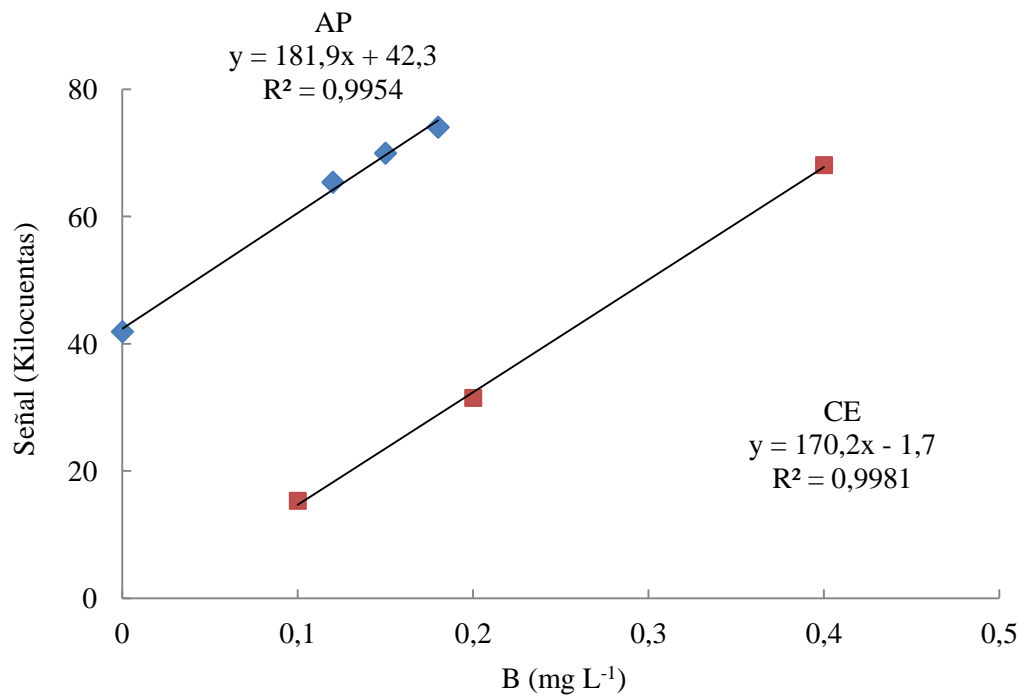


Figura IV.17. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de boro.

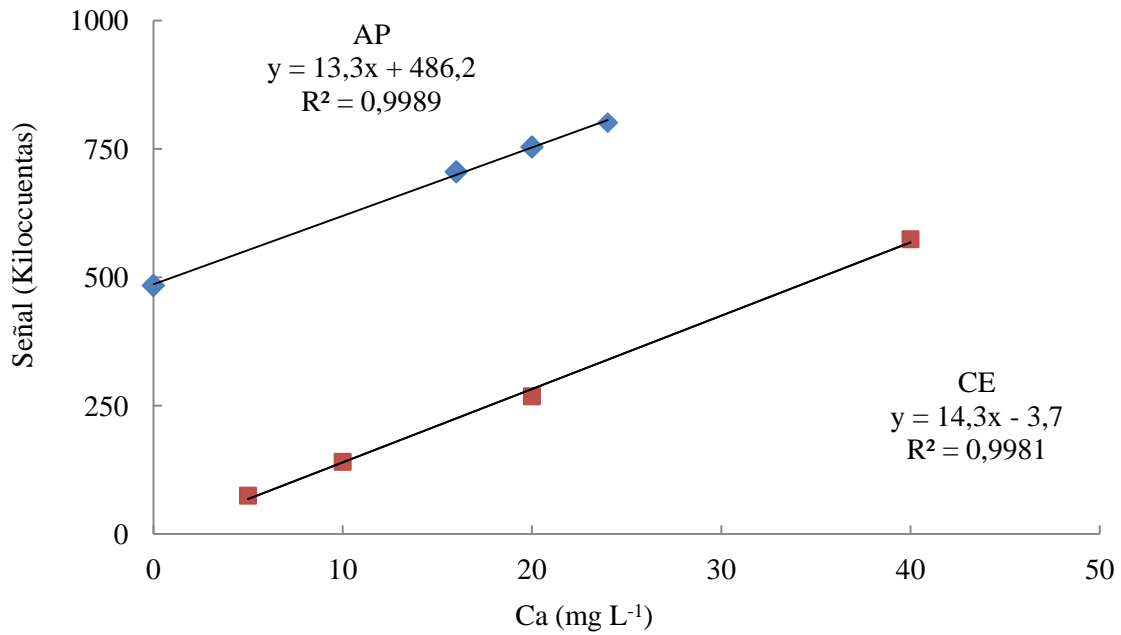


Figura IV.18. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de calcio.

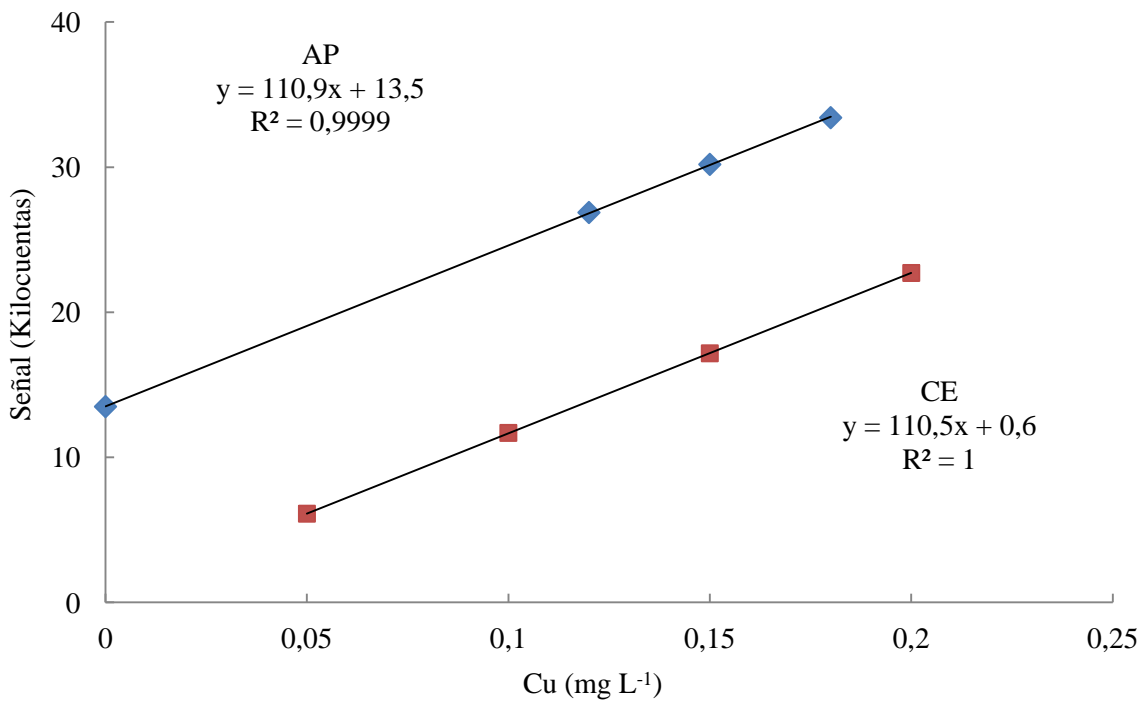


Figura IV.19. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de cobre.

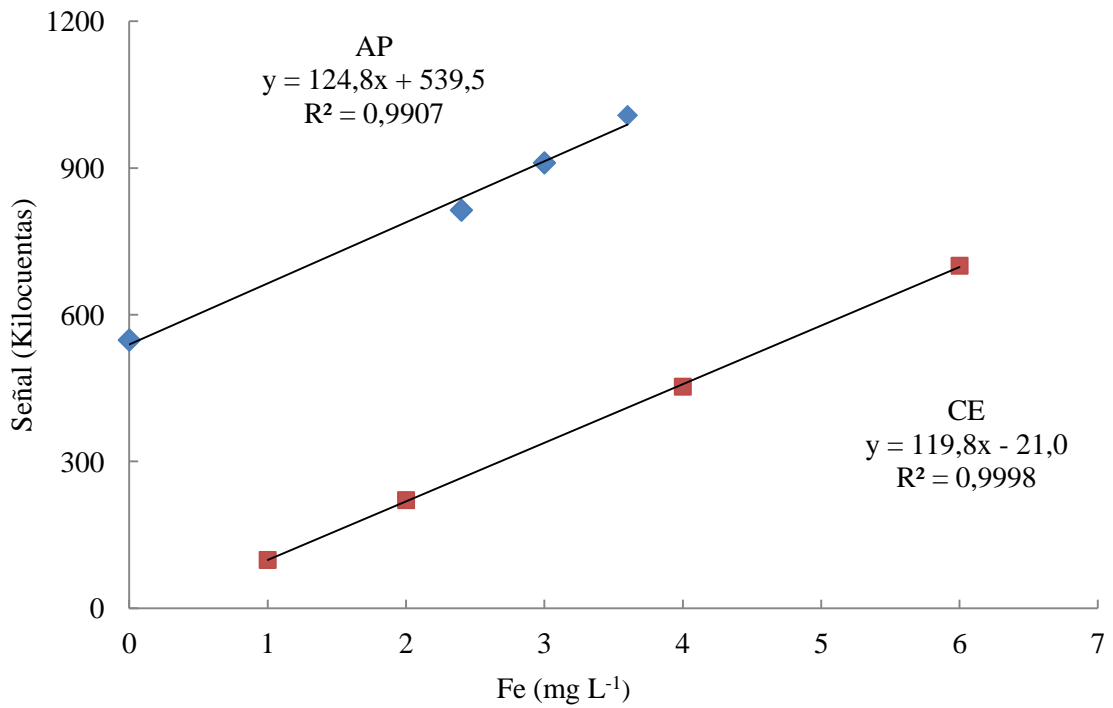


Figura IV.20. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de hierro.

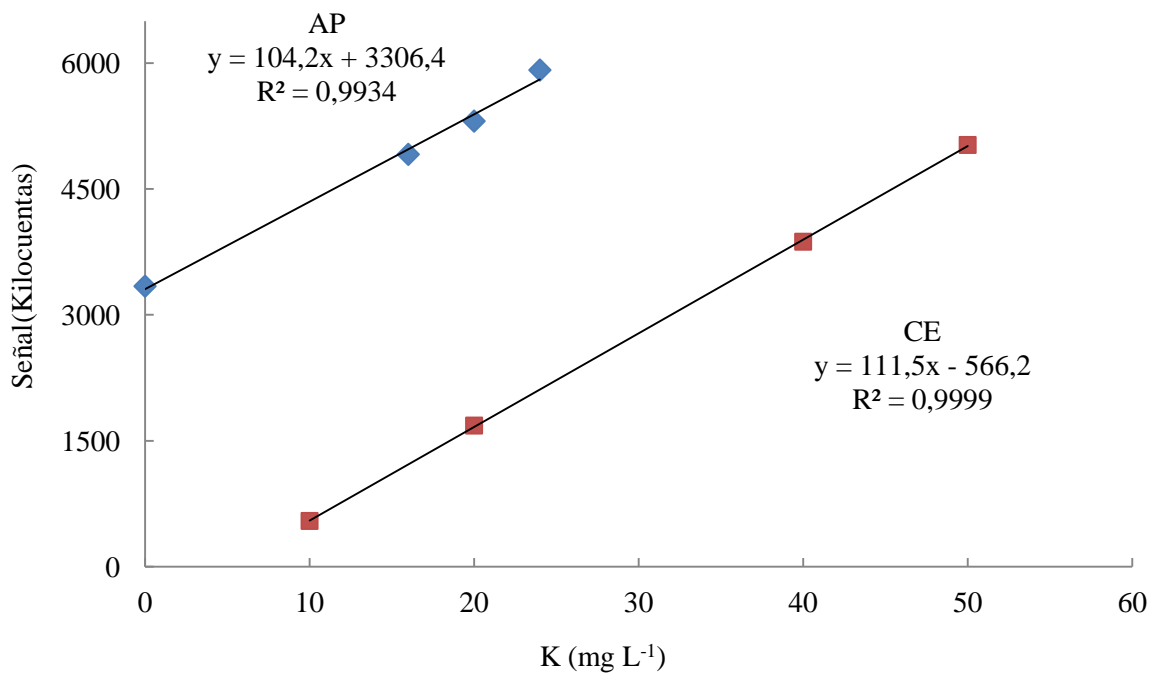


Figura IV.21. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de potasio.

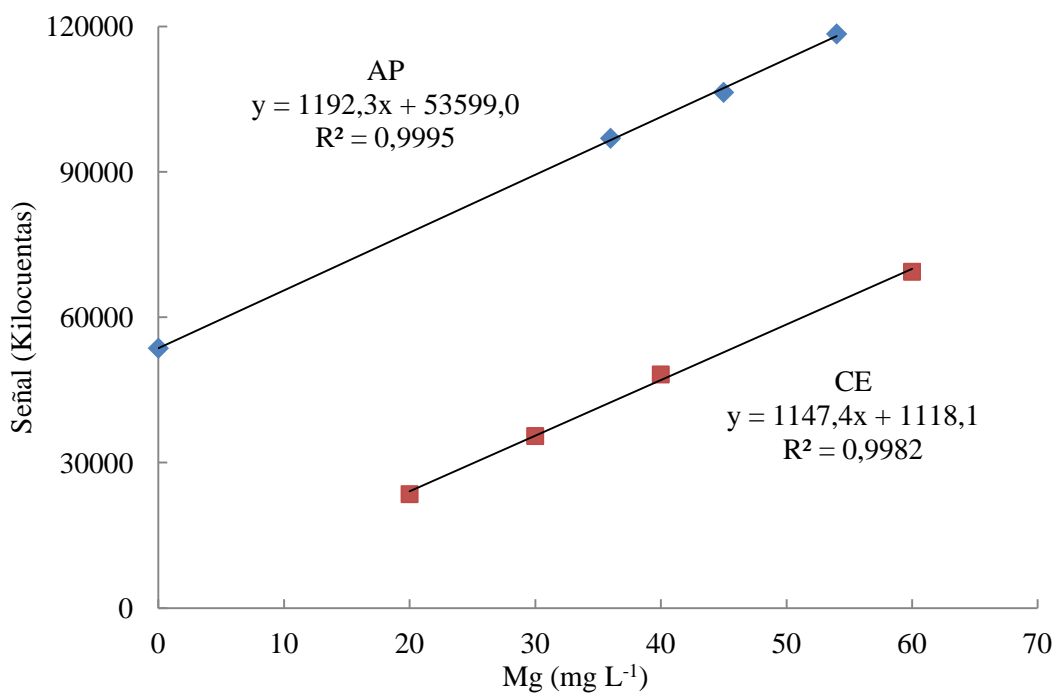


Figura IV.22. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de magnesio.

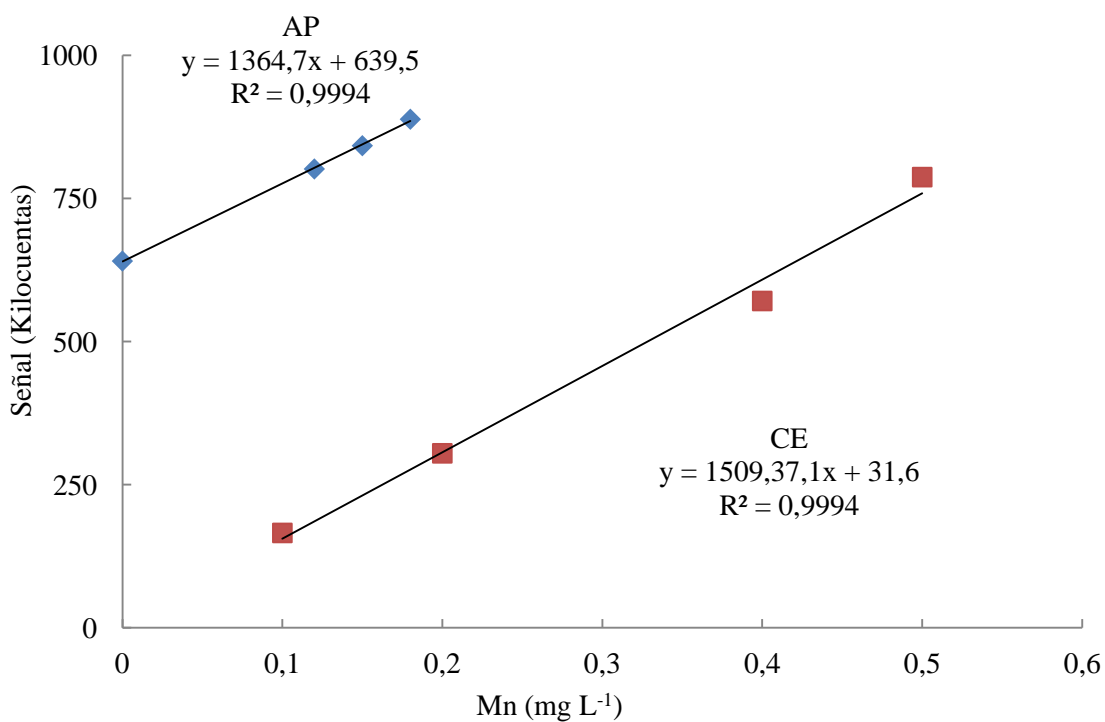


Figura IV.23. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de manganeso.

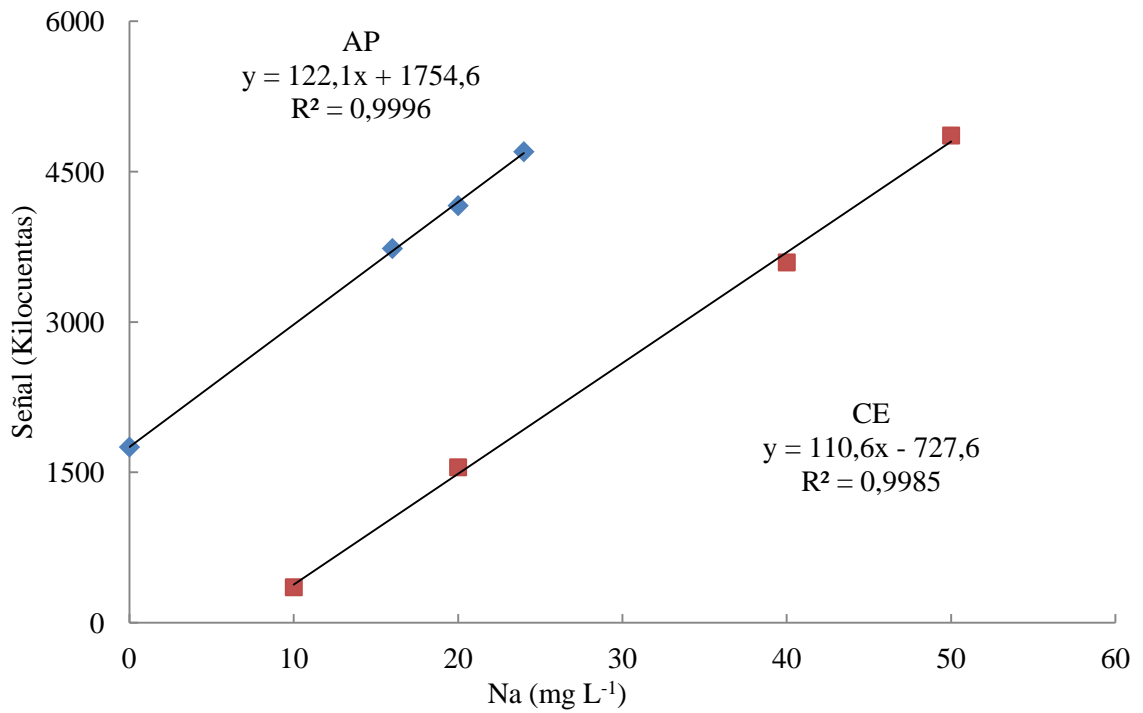


Figura IV.24. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de sodio.

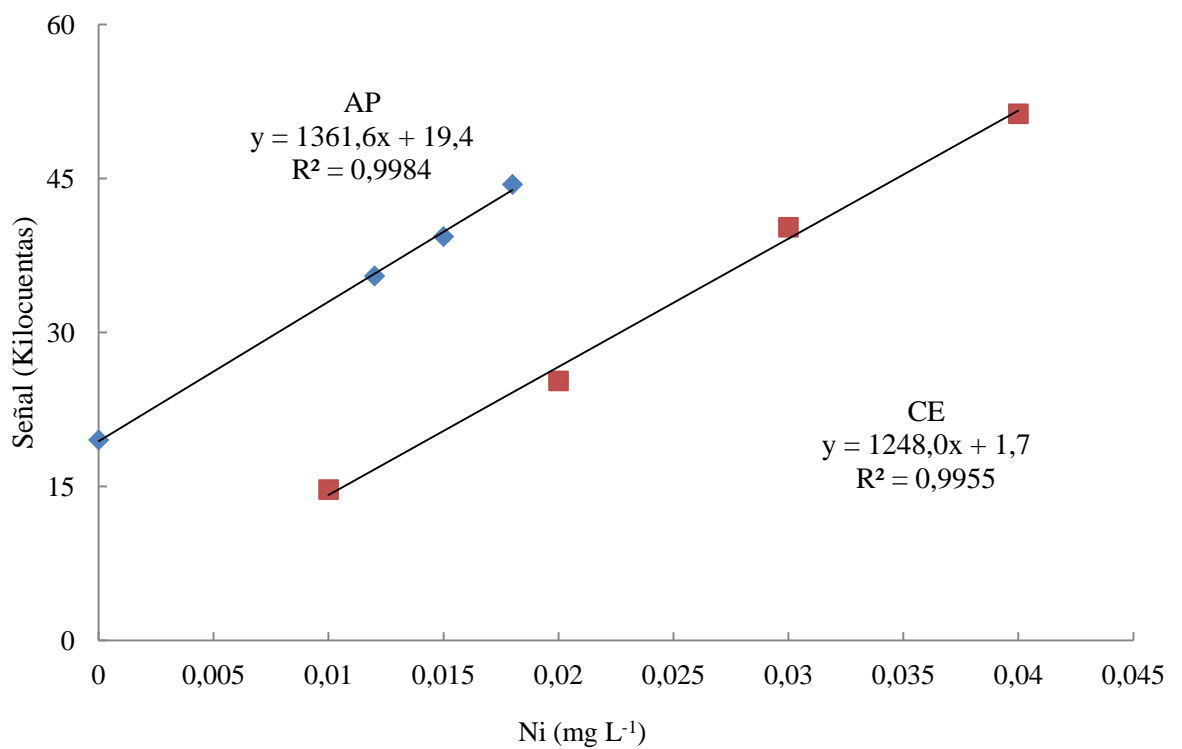


Figura IV.25. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de níquel.

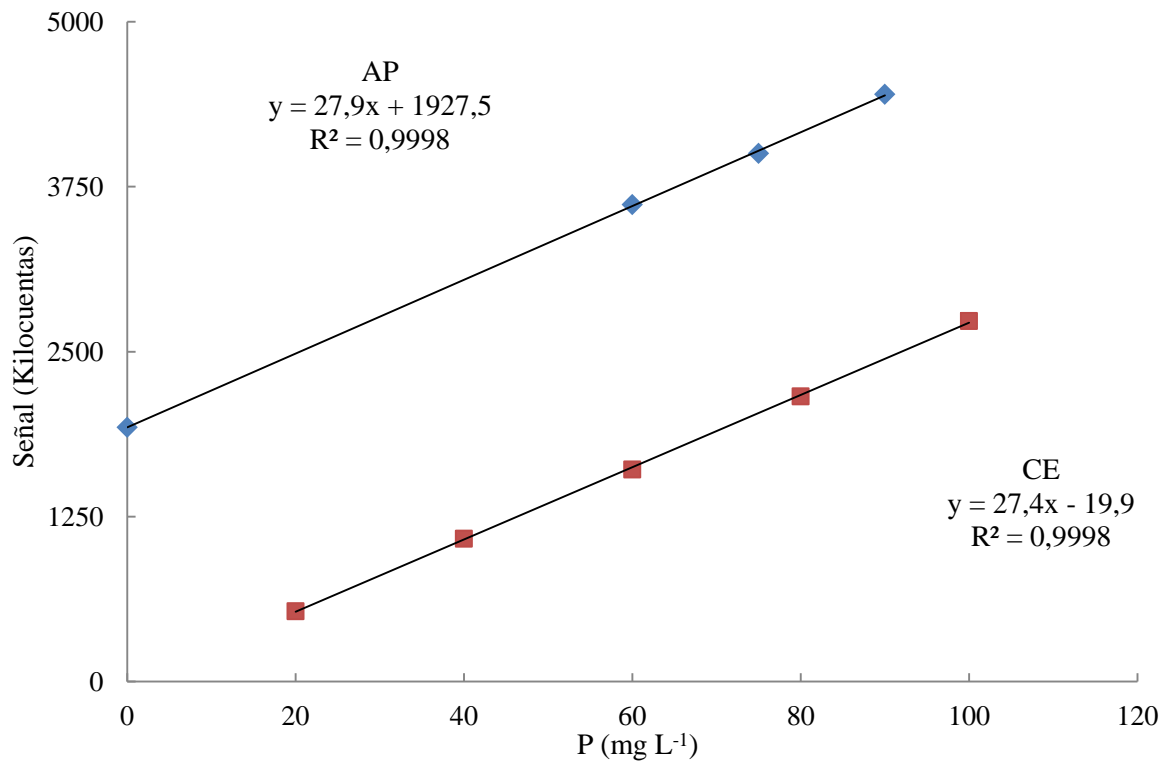


Figura IV.26. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de fósforo.

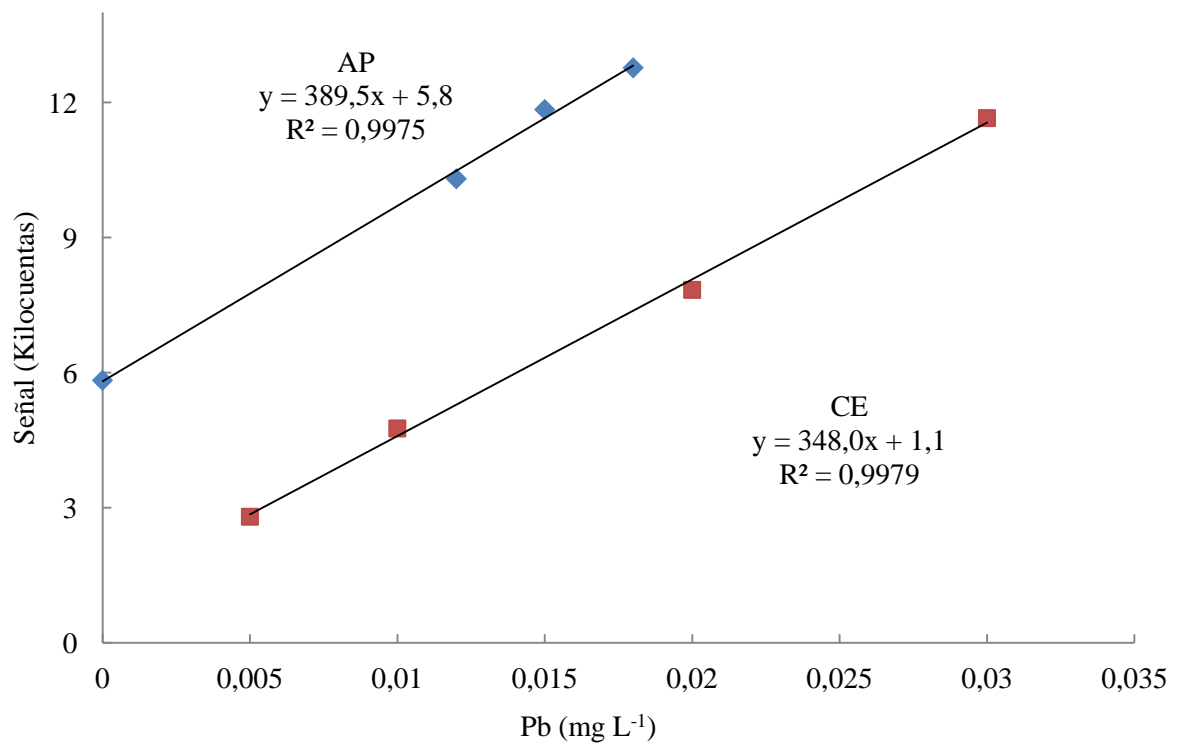


Figura IV.27. Recta de calibrado externo (CE ■) y adición patrón (AP ◆) de plomo.

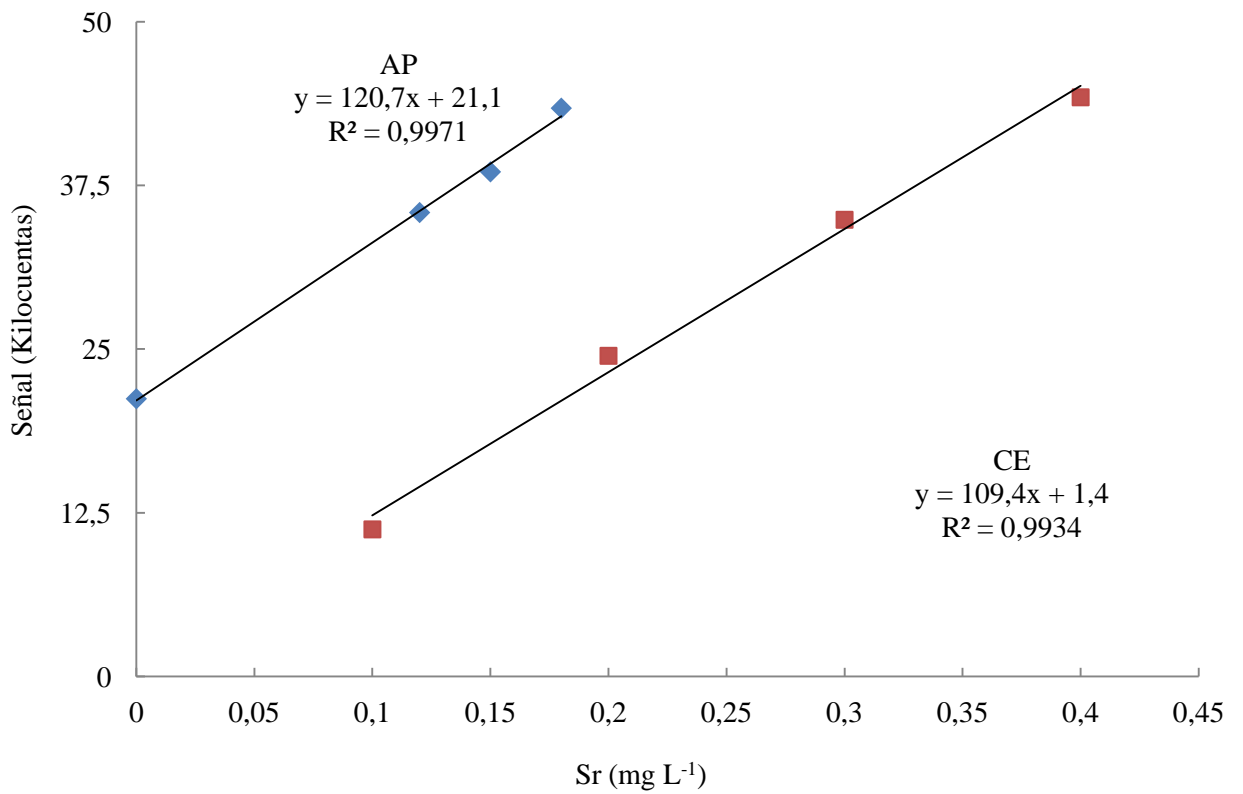


Figura IV.28. Recta de calibrado externo (CE ■) y adición patrón (AP◆) de estroncio.

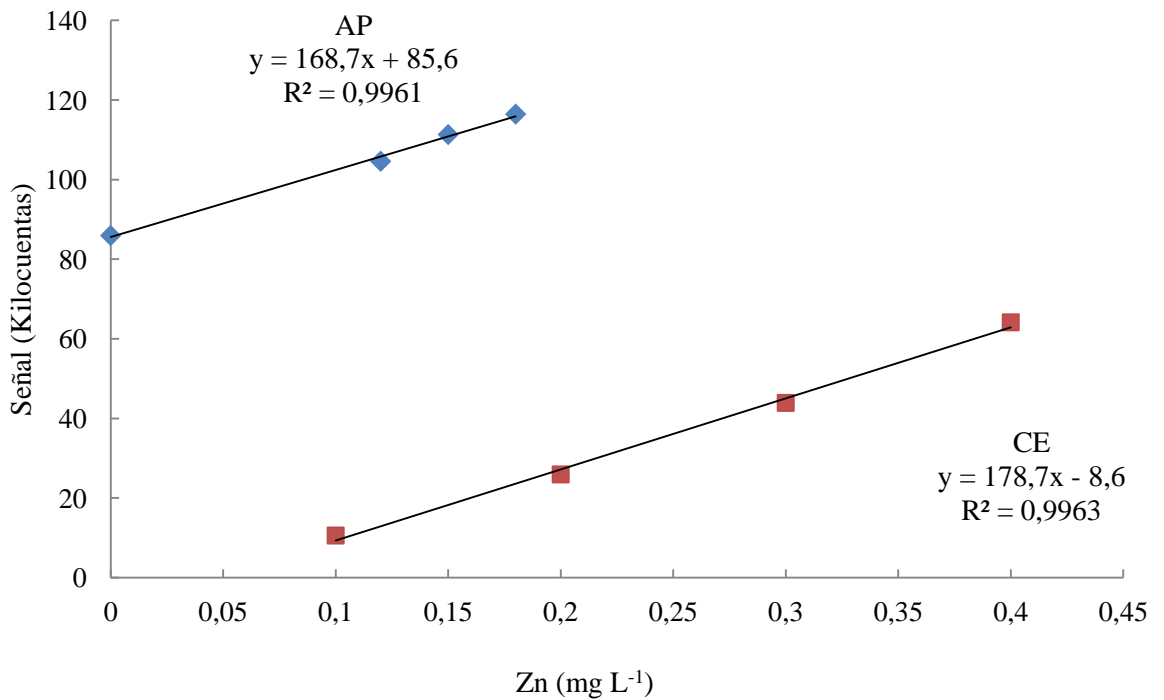


Figura IV.29. Recta de calibrado externo (CE ■) y adición patrón (AP◆) de zinc.

IV.1.3. Veracidad

A partir de los datos incluidos en la tabla IV.3. para el método M3 se comprueba la veracidad del método para la determinación de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn. Se analizaron muestras fortificadas en tres niveles de adición, realizando tres repeticiones en cada nivel. Como se observa, los intervalos de los valores de recuperación obtenidos contienen el 100% para todos los elementos analizados, confirmándose que el método es veraz.

IV.1.4. Precisión

La precisión del método se estudió en condiciones de repetitividad y de precisión intermedia. La repetitividad se refiere a la precisión obtenida al realizar medidas repetidas en un día sin variar las condiciones operacionales. La precisión intermedia se refiere al grado de concordancia de los resultados obtenidos entre sí cuando se realizan las medidas en días distintos cambiando alguna condición experimental.

La precisión se puede estudiar a partir de un análisis de varianza (ANOVA). Para ello realizamos n replicados durante h días en condiciones de precisión intermedia. La varianza de repetitividad o error puro (varianza dentro el grupo), S_{PE}^2 , vendrá dada por la ecuación 4.14,

$$S_{PE}^2 = \frac{\sum_i^h s_i^2}{h} = \frac{\sum_i^h \sum_j^n (X_{ij} - \bar{X}_i)^2}{h(n-1)} \quad (4.14)$$

mientras que la debida al factor o varianza entre días (varianza entre grupos), S_{Factor}^2 , vendrá dada por la ecuación 4.15,

$$S_{Factor}^2 = \frac{n \sum_i^h (\bar{X}_i - \bar{X})^2}{h-1} \quad (4.15)$$

Como en la ecuación 4.15 están incluidas las medidas de cada grupo, se estaría considerando el efecto de la repetitividad dentro del efecto del factor. Por eso, se debe

calcular una varianza neta entre grupos que solo depende de la varianza entre las medidas de los grupos y la media global, tal y como se muestra en la ecuación 4.16.

$$S_B^2 = \frac{(S_{Factor}^2 - S_{PE}^2)}{n} \quad (4.16)$$

La precisión intermediase calcula finalmente como la suma de la repetitividad (error puro) y la varianza neta debida al factor (neta entre grupos) según la ecuación. 4.17 [9].

$$S_R^2 = S_{PE}^2 + S_B^2 \quad (4.17)$$

Horwitz dedujo una expresión (ecuación 4.18) para predecir el valor esperado para la desviación estándar relativa a partir de la concentración de analito, en tanto por 1.

$$RSD_H = 2^{(1-0,5\log c)} \quad (4.18)$$

El valor de RSD obtenido se compara con el predicho por Horwitz (RSD_H) mediante el parámetro Horrat [10]:

$$Horrat = \frac{RSD_R}{RSD_H} \quad (4.19)$$

Valores iguales a uno muestran una correspondencia adecuada entre el valor predicho y el obtenido, mientras que valores mayores que uno y particularmente mayores que dos, indican precisiones no aceptables [11].

Los valores de repetitividad y precisión intermedia se dan como desviación estándar relativa y se presentan en la tabla IV.5. junto con el parámetro de Horrat. Los valores de repetitividad varían entre 0,72 y 10,21 %. Los elementos Zn, Ni y Mn tienen los valores más elevados, mientras que P y Al los tienen más bajos. En el caso de la precisión intermedia, el rango de valores oscila entre 1,17 y 11,69 %. Zn, Ni y Mn son los elementos que presentan los valores de precisión intermedia más altos, y Al y P tienen los menores valores.

Tabla IV.5. Repetitividad y precisión intermedia, dados como desviación estándar relativa junto con el parámetro Horrat.

Elemento	Repetitividad	Precisión intermedia	Horrat
Al	0,88	1,86	0,14
B	6,37	8,38	0,43
Ca	2,44	3,50	0,37
Cu	3,07	3,97	0,18
Fe	2,25	3,89	0,29
K	2,10	2,79	0,29
Mg	2,04	3,16	0,35
Mn	8,26	11,43	0,63
Na	4,34	4,69	0,47
Ni	7,93	10,00	0,34
P	0,72	1,17	0,14
Pb	4,19	4,74	0,16
Sr	1,57	8,30	0,40
Zn	10,21	11,69	0,66

Los valores de Horrat calculados varían entre 0,14 y 0,66. Al, Cu, P y Pb tienen los valores más pequeños, siendo 0,14, 0,18, 0,14 y 0,16, respectivamente. Para Mn y Zn los valores de Horrat son más altos, 0,63 y 0,66, respectivamente. Como puede apreciarse, el valor del Horrat es inferior a 1 para todos los elementos determinados, por lo tanto se puede considerar que la precisión del método es aceptable, tanto en condiciones de repetitividad, como en condiciones de precisión intermedia.

IV.1.5. Calibración y linealidad

Teniendo en cuenta la ausencia de efecto matriz, se utilizó el método de calibrado externo para la cuantificación de los elementos en las muestras. Los parámetros de las rectas de calibrado, obtenidos utilizando la intensidad de emisión en kilocuentas como señal analítica y la concentración en mgL^{-1} , se presentan en la tabla IV.6.

La linealidad se establece a partir de la pendiente de la recta de calibrado y su desviación estándar, según la ecuación 4.20 [12].

$$\text{Linealidad} (\%) = \left(1 - \frac{S_b}{b}\right) \cdot 100 \quad (4.20)$$

En la tabla IV.6.se muestran los valores de linealidad encontrados para los elementos estudiados. Como puede observarse las linealidades oscilan en el intervalo de 95-99,1.

IV.1.6. Límite de detección y límite de cuantificación

El límite de detección (LOD) se define como la cantidad de analito presente que da lugar a la menor señal analítica que puede detectarse con certeza razonable (Y_{LOD})[13]. Esta señal puede estimarse a partir de medidas repetidas del blanco de muestra a partir de la expresión:

$$Y_{LOD} = \bar{Y}_0 \pm 3s_{\bar{Y}_0} \quad (4.21)$$

donde \bar{Y}_0 es la media de las señales medidas para el blanco y $s_{\bar{Y}_0}$ su desviación estándar [13]. A partir de los parámetros de la recta de calibrado se puede calcular la concentración correspondiente como:

$$X_{LOD} = \frac{Y_{LOD} - a}{b} \quad (4.22)$$

Tabla IV.6. Rectas de calibrado y linealidad para los distintos elementos determinados

Elemento	Rango de concentración (mgL ⁻¹)	b ± s _b	a ± s _a	R ²	Linealidad (%)
Al	0-10	103,3± 23	61,6± 157	0,999	98,8
B	0,2-1	152,4± 8	3,9± 5	0,999	96,8
Ca	20-100	13,1 + 1	18,9± 53	0,996	99,1
Cu	0,2-1	109,1± 3	-0,34 ± 2	0,999	99,0
Fe	0-10	133,3± 5	- 71,5± 34	0,997	98,9
K	20-100	85,8± 4	640,5± 237	0,999	98,6
Mg	20-100	1159,0 + 15	786,5± 977	0,999	95,3
Mn	0,2-1	1516,5± 61	-33,6± 41	0,999	97,2
Na	20-100	123,3± 4	- 1161,2± 240	0,999	98,4
Ni	0,02-0,1	1401,6± 43	- 3,1± 2	0,999	98,7
P	20-100	27,4 + 0,22	- 20,7± 15	0,999	99,1
Pb	0,02-0,1	354,8± 11	1,6± 0,3	0,999	96,9
Sr	0,2-1	117,2± 3	-0,31 ± 1	0,999	98,4
Zn	0,2-1	176,8± 6	-9,1± 4	0,998	98,3

b: pendiente de la recta de calibrados. s_b: desviación estándar de la pendiente. a: ordenada en el origen de la recta de calibrado. s_a: desviación estándar de ordenada en el origen. R²: coeficiente de determinación.

Puesto que la señal del blanco debe coincidir con la ordenada en el origen (a), la expresión anterior se reduce a:

$$X_{LOD} = \frac{3s_{\bar{Y}_0}}{b} \quad (4.23)$$

El límite de cuantificación (LOQ) se define como la cantidad de analito presente derivada de la menor señal analítica que puede ser medida con exactitud. Se calcula como:

$$X_{LOQ} = \frac{10s_{\bar{Y}_0}}{b} \quad (4.24)$$

La tabla IV.7.incluye los valores de LOD y LOQ obtenidos. Los límites de detección, están comprendidos entre 0,001 y 0,064 mgL⁻¹. Salvo en los casos del B, K y P, todos los valores están por debajo de 0,010 mgL⁻¹.Los mayores límites de

cuantificación encontrados son el B y el K, que son 0,210 y 0,15 mgL⁻¹, respectivamente. El resto de elementos presentan límites de cuantificación que oscilan desde 0,003 mgL⁻¹, como es el caso del Fe, Mn, Sr, Zn o Mg, hasta 0,020 mgL⁻¹ para el Ni.

Tabla IV.7. LOD y LOQ para los elementos determinados.

Elemento	LOD (mgL ⁻¹)	LOQ (mgL ⁻¹)
Al	0,002	0,007
B	0,064	0,21
Ca	0,002	0,007
Cu	0,009	0,03
Fe	0,001	0,003
K	0,045	0,150
Mg	0,001	0,003
Mn	0,001	0,003
Na	0,008	0,027
Ni	0,006	0,02
P	0,011	0,037
Pb	0,008	0,03
Sr	0,001	0,003
Zn	0,001	0,003

IV.1.7. Determinación de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn en muestras de pimentón.

Los valores determinados se refieren a base seca. Para ello, se ha determinado la humedad de las muestras de acuerdo con el procedimiento explicado en el apartado III.2.1. En todos los casos encontramos que los valores de humedad calculados para las muestras están por debajo del máximo exigido por la legislación vigente para los pimentones de Extremadura y Murcia.

Siguiendo el procedimiento descrito se ha determinado el contenido de Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr y Zn, en muestras de pimentón. Los resultados se han incluido en la tabla del anexo 1 A1.1.

Como se puede observar, el elemento más abundante es el K, con concentraciones del orden de 15000 mg kg⁻¹ o superiores en muchas muestras. El P es el elemento con mayor concentración después del K, y los valores están en torno a 3500 mg kg⁻¹. Ca y Mg están presentes en concentraciones similares y del orden de 2000 mg kg⁻¹. Las concentraciones de Na oscilan alrededor de 500 mg kg⁻¹, que son bastante inferiores a las de K, y por lo tanto se puede considerar que el pimentón es un alimento rico en K [14], lo cual es importante para mantener el equilibrio Na/K en el cuerpo. Una parte de la energía del cuerpo se utiliza para mantener las concentraciones de K y Na en una relación dada, ya que dicho equilibrio es importante para mantener la función del corazón, transmisión nerviosa y contracción muscular. Otros elementos que se encuentran en concentraciones del orden de 200 mg kg⁻¹ son el Fe y el Al, siendo en general mayores los valores para el Fe. Zn y Mn tienen concentraciones inferiores a 30 mg kg⁻¹, aunque en algunas muestras los valores están por encima de este valor. B, Cu y Sr presentan niveles inferiores y próximos a 15 mg kg⁻¹. Como se puede observar, los elementos minoritarios son Pb y Ni. Los contenidos medios de estos elementos son 0,42 y 0,97 mg kg⁻¹ respectivamente.

Las muestras que presentan menos concentración de Na son las de Extremadura, 346,31 mg kg⁻¹, mientras de las muestras de Murcia presentan valores de concentración para el Na de 502,55 mg kg⁻¹. En los demás casos las concentraciones de Ca, Mg, P y K son muy similares entre los dos tipos. Para elementos como Al, B, Cu, Mn, Sr o Zn encontramos que las concentraciones son similares, siendo los valores más altos en Extremadura para los elementos Al, Mn, Zn y B y más bajos en Cu, Sr y Fe. Estas tendencias se analizarán con más detalle para llevar a cabo un estudio de diferenciación del origen del pimentón.

Estos resultados son acordes con los publicados por otros autores como Brunner et al. [15] que determinaron el contenido de metales en muestras de pimentón húngaro y Filgueiras et al. [16] que determinan la concentración de algunos metales presentes en muestras de pimentón.

IV.2. Diferenciación de origen geográfico de pimentón según su perfil metálico

IV.2.1. Análisis estadístico básico

Las muestras de pimentón estudiadas son de tres tipos diferentes: dulces, picantes y agridulces. Para realizar el estudio de diferenciación geográfica se comprueba previamente si debe realizarse considerando los tipos por separado o tomando todas las muestras en conjunto. Para ello, se comprueba si existen diferencias significativas entre los tipos de pimentón. Teniendo en cuenta los valores incluidos en la tabla IV.8. se aprecia que hay variaciones importantes en las concentraciones de los elementos determinados en los grupos de muestras correspondientes a cada DOP. No obstante, puede observarse, si se comparan los orígenes MU y EX sin atender al tipo de pimentón, que existen ciertas diferencias. En el caso del Al, Ca, Fe, Mg, Na y Sr, encontramos valores máximos mayores para el origen de MU que para el de EX. En concreto, para elementos como Na, los valores de concentración en muestras de la DOP MU oscilan entre 168,2 y 947,6 mg kg⁻¹, mientras que para las muestras de la DOP EX los valores se encuentran dentro del rango 154,6-559,9 mg kg⁻¹. Para el Al los valores máximos son 245,7 mg kg⁻¹ para EX y 276,3 mg kg⁻¹ para MU. Al considerar los valores para el Cu, se comprueba que las diferencias entre los valores máximos son menos marcadas, ya que para EX el valor máximo es 22,6, mientras que para MU es 20,6. En los casos del P y Zn las variaciones tienen una tendencia similar a la comentada para el Cu. El P tiene concentraciones máximas de 4266,1 y 4226,4 mg kg⁻¹ en las DOP EX y MU, respectivamente, mientras que para el Zn los valores son 69,6 y 69,0 mg kg⁻¹, respectivamente. Fe, Mg, Na, Sr presentan unos valores que tienen variaciones más similares a las comentadas para Al y Na. En el caso de B, K, Mn, Ni, y Pb, la tendencia es inversa, ya que los valores máximos mayores corresponden a las muestras de la DOP EX. Por ejemplo, la concentración máxima de K para muestras EX es 23416,4 mg kg⁻¹, mientras que el valor en el caso de las muestras MU es 20157,8 mg kg⁻¹. Considerando las medianas se aprecia que los valores son mayores para los elementos Al, B, K, Mg, Mn, P, Pb y Zn en la DOP EX, y menores para Cu, Fe, Na y Sr en la DOP MU. El elemento Ca presenta valores similares en ambas DOP.

Tabla IV.8. Mediana, mínimo y máximo (mg kg^{-1}).

Elemento	MURCIA	EXTREMADURA
Al	118,1 (26,4-276,3)	123,8 (79,6-245,7)
B	12,2 (0,263-23,6)	13,4 (5,8-27,9)
Ca	1911,4 (1011,8-2789,8)	1911,9 (832,9-2465,1)
Cu	11,6 (9,11-22,6)	10,3 (6,98-20,6)
Fe	200,6 (85,3-418,5)	178,0 (83,8-295,4)
K	14685,2 (7691,0-20157,8)	14927,7 (10318,2-23416,4)
Mg	1999,1 (1642,3-2717,6)	2179,2 (1483,4-2460,9)
Mn	14,8 (9,05-53,1)	27,6 (15,2-97,6)
Na	502,6 (168,2-947,6)	346,3 (154,6-559,9)
Ni	0,93 (0,543-1,56)	1,01 (0,173-3,40)
P	3495,0 (2685,5-4226,4)	3591,0 (2498,7-4266,1)
Pb	0,372 (0,102-11,2)	0,471 (0,03-21,3)
Sr	16,10 (8,49-38,5)	9,17 (5,48-29,1)
Zn	31,7 (14,2-69,0)	33,9 (18,3-69,6)

Valores mínimos y máximos entre paréntesis.

En la tabla IV.9. se han incluido los valores de mediana, máximo y mínimo para los grupos de muestras estudiadas, diferenciando el origen geográfico y los tipos. Dentro del origen MU, encontramos que en todos los elementos existen diferencias entre tipos para los valores de la mediana. En el caso de los pimentones de EX se encuentra una tendencia similar, excepto para el caso del Ni en el que los valores de la mediana para los tres tipos, dulce, picante y agridulce, son prácticamente iguales.

Si comparamos valores de mediana de muestras pimentón picante entre los orígenes MU y EX, encontramos que las diferencias no son muy acusadas, siendo mayores los valores para Al, Ca, Cu, Fe, K, Mg y Na en el origen MU, mientras que B, Mn, Ni, P, Pb, Sr y Zn son mayores para el origen EX. En el caso de los dulces, existen diferencias importantes entre los valores de las medianas, siendo mayores en las muestra de origen MU que en EX. Así mismo, en las muestras de tipo agrídulce se observan valores más elevados para EX.

Para visualizar de una forma más clara las tendencias de los datos, se han representado en los diagramas de caja y bigotes de las figuras IV.30.- IV.35., la mediana y los datos incluidos en el primer cuartil para los elementos determinados. Como primera aproximación, los diagramas de caja y bigotes se han realizado atendiendo solamente al origen geográfico.

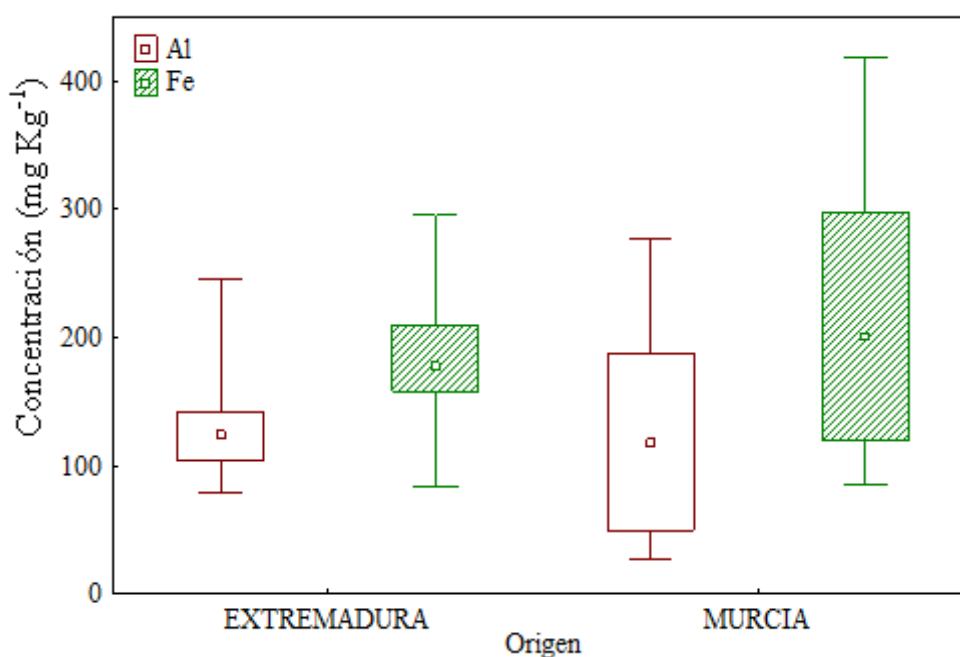


Figura IV.30. Diagrama de caja y bigotes para Al y Fe.

En el caso del Al los valores de las medianas de los orígenes MU y EX son 118,1 y 123,8 mg kg⁻¹ respectivamente y no muestran una gran diferencia. Para el Fe la situación es similar, ya que los valores de mediana de los orígenes MU y EX son 200,6y

178,0 mg kg⁻¹, respectivamente, aunque se observa una mayor dispersión en el origen MU.

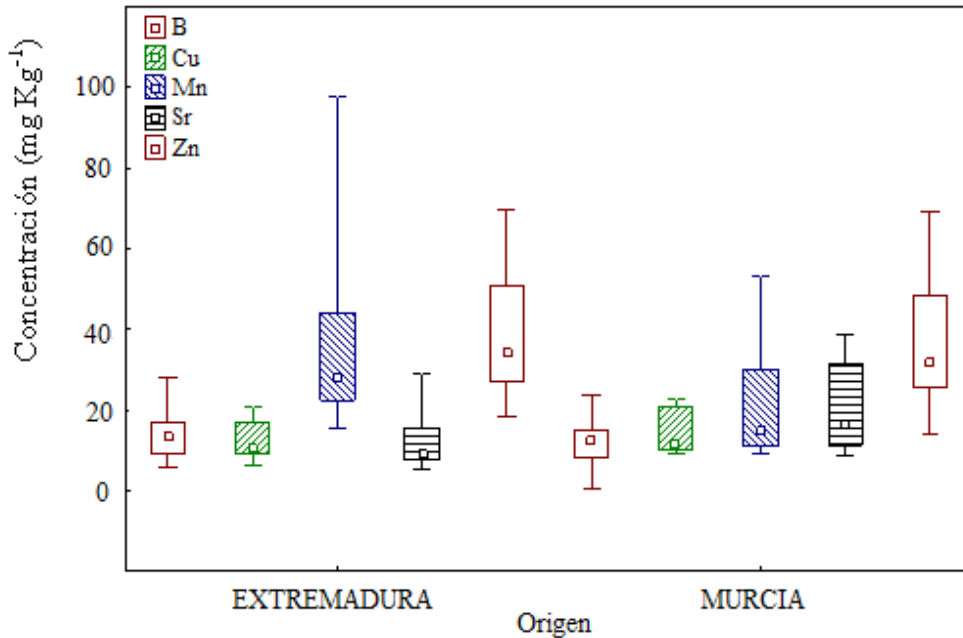


Figura IV.31. Diagrama de caja y bigotes para B, Cu, Mn, Sr y Zn.

Considerando los elementos B, Cu y Zn, tal y como se observa en la figura IV.31., los valores de las medianas para ambos orígenes son similares. En el caso del B, para el origen MU es 12,2 y para EX 13,4 mg kg⁻¹. Los valores para Cu son 11,6 y 10,3 mg kg⁻¹ para MU y EX, respectivamente. Para el Zn, los valores de las medianas son 31,7 y 33,9 mg kg⁻¹, para MU y EX, respectivamente. El recorrido es similar para estos tres elementos en los dos orígenes. Para el Sr y Mn se aprecia una mayor diferencia en los valores de las medianas, que para el Sr son 16,10 y 9,17 mg kg⁻¹ para MU y EX, respectivamente, mientras que para el Mn son 14,8 y 27,6 mg kg⁻¹, respectivamente. La distribución de los datos en los distintos cuartiles varía en los dos orígenes para Sr y Mn, siendo mayor para el Sr en el origen MU, y menor para el Mn en este origen. El valor máximo del Mn en el origen EX es especialmente elevado.

Tabla IV.9. Valores de la mediana, mínimo y máximo mg kg⁻¹ del contenido metálico en muestras de pimentón.

	MURCIA			EXTREMADURA		
	Dulce	Picante	Agridulce	Dulce	Picante	Agridulce
Al	139,6 (94,8-276,3)	136,3 (43,0-352,6)	42,15 (26,4-136,8)	132,0 (79,6-245,7)	117,2 (83,5-160,1)	128,0 (91,1-227,3)
B	14,2 (7,38-20,8)	12,2 (0,260-22,6)	7,92 (5,46-23,6)	14,8 (7,28-27,9)	13,5 (7,01-23,5)	10,5 (5,77-23,8)
Ca	1946,2 (1498,0-2442,0)	2191,1 (1011,8-2789,8)	1493,8 (1030,6-2174,1)	1894,800 (1062,6-2465,1)	2043,8 (832,9-2436,8)	1862,1 (1112,6-2421,7)
Cu	20,3 (9,25-22,2)	12,9 (9,11-22,6)	11,1 (9,21-22,6)	11,6 (5,98-20,0)	12,7 (7,4-20,6)	9,84 (6,49-20,6)
Fe	228,7 (182,5-402,9)	200,6 (123,3-418,5)	105,9 (85,3-264,5)	178 (83,8-295,4)	167,8 (123,1-235,7)	183,1 (150,4-246,9)
K	15687 (9169,0-20157,8)	14537,8 (9226,0-19765,2)	14387,3 (7691,0-17547,5)	14734,0 (11174,7-17581,2)	14261,4 (10639,6-23416,5)	15205,8 (10318,2-20213,2)
Mg	2011,8 (2134,2-2011,8)	2256,5 (1703,4-2717,6)	1785,4 (1642,3-2190,3)	2188,6 (1483,4-2460,9)	2174,4 (1866,5-2389,9)	2194,3 (1950,6-2345,6)
Mn	24,9 (10,7-35,8)	20,5 (12,3-53,1)	10,3 (9,05-27,7)	26,9 (19,7-97,6)	26,1 (15,3-55,6)	28,7 (19,1-68,5)
Na	509,1 (418,9-880,6)	440,6 (168,2-783,7)	545,6 (277,8-947,6)	327,6 (154,6-531,1)	333,3 (255,4-496,6)	377,3 (21,0-559,9)
Ni	0,930 (0,810-1,43)	1,06 (0,830-1,56)	0,610 (0,540-0,880)	1,00 (0,170-2,94)	1,01 (0,520-2,1)	1,02 (0,490-3,40)
P	3396,2 (3080,9-3575,2)	3555,2 (2685,5-4092,6)	3643,8 (3154,4-4226,3)	3613,2 (2498,7-4266,1)	3568 (3169,5-3950,1)	3619,3 (3263,0-4139,9)
Pb	0,490 (0,03-10,7)	0,400 (0,09-11,2)	0,100 (0,01-9,08)	0,400 (0,03-19,0)	1,12 (0,100-16,8)	0,420 (0,04-21,3)
Sr	29,8 (9,59-38,5)	0,400 (8,5-38,4)	12,2 (9,04-32,1)	8,71 (5,68-29,1)	9,40 (5,48-21,3)	9,18 (6,62-16,3)
Zn	43,3 (14,9-62,2)	15,1 (14,2-60,8)	30,5 (23,9-69,0)	35,7 (14,2-62,7)	32,0 (18,9-62,2)	33,4 (18,3-60,5)

Valores mínimos y máximos entre paréntesis.

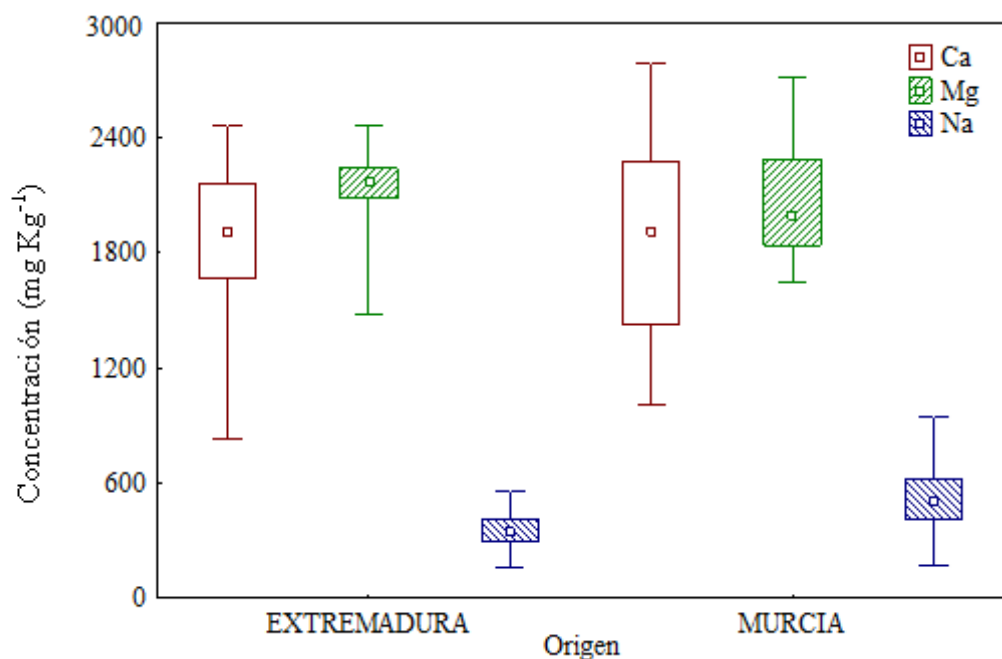


Figura IV.32. Diagrama de caja y bigotes para Ca, Mg y Na.

Para el caso del Ca, en la figura IV.32. se puede ver como los valores de las medianas son similares, 1911,4 y 1911,9 mg kg⁻¹ para MU y EX, respectivamente. El rango en el que oscilan los valores del primer cuartil es mayor para el origen MU. Los recorridos para ambos orígenes son similares, aunque el valor mínimo del Ca en el origen EX es bastante bajo.

Considerando los valores para el Mg se aprecia que las medianas no son tan similares siendo 1999,1 y 2179,2 mg kg⁻¹ para MU y EX, respectivamente. La dispersión de datos en el primer cuartil es mayor para el origen MU, aunque los recorridos son similares.

En el caso del Na, se puede observar que el valor de la mediana para el origen EX, 346,3 mg kg⁻¹, es menor que en el caso del origen MU, que tiene un valor igual a 502,6 mg kg⁻¹. La dispersión de los valores en el primer cuartil es mayor en el origen MU, e igual ocurre con el recorrido, que es mayor en MU.

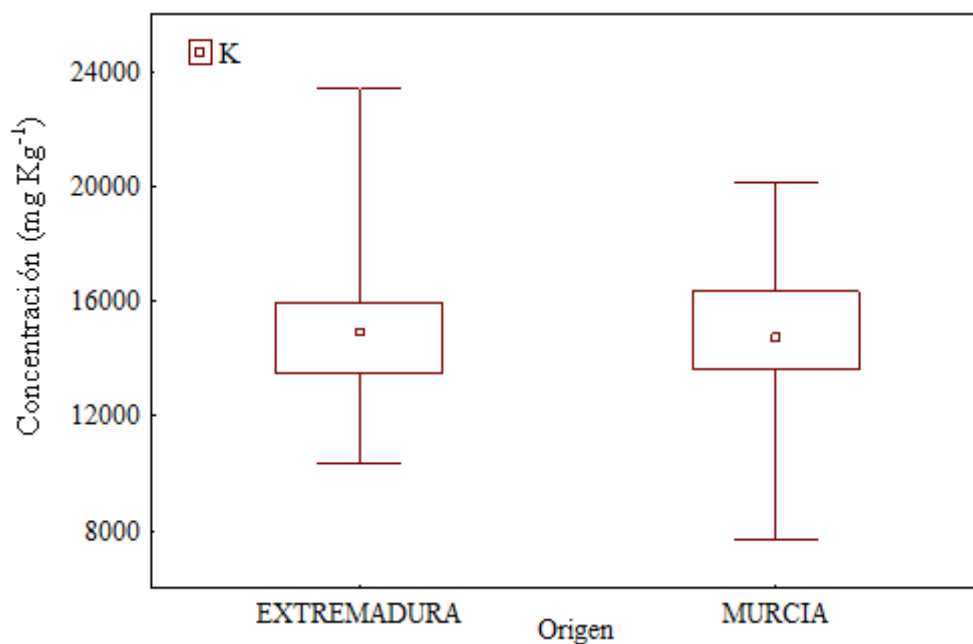


Figura IV.33. Diagrama de caja y bigotes para K.

La figura IV.33 muestra los diagramas de caja y bigotes de los datos correspondientes al elemento K para los dos orígenes geográficos. En ambos casos, los valores de las medianas son bastante parecidos, 14685,2 y 14927,7 mg kg⁻¹, para las DOP MU y EX, respectivamente. Las dispersiones en el primer cuartil y el recorrido son similares en ambos orígenes, aunque los valores mínimo y máximo para el origen MU son menores.

En la figura IV.34 se muestra el diagrama de caja y bigotes para el P. Se puede apreciar que el valor de la mediana para el origen EX, 3591,0 mg kg⁻¹, es algo mayor que el correspondiente al origen MU, 3495 mg kg⁻¹. Hay una mayor dispersión de los datos en el primer cuartil para MU, mientras que el recorrido es mayor en EX.

En la figura IV.35., se han incluido los diagramas de caja y bigotes Pb y Ni. Se observa que los valores de la mediana del Pb, 0,372 y 0,471 mg kg⁻¹, para MU y EX, respectivamente, son parecidos en ambos casos orígenes. En el caso del Ni los valores 0,93 y 1,01 mg kg⁻¹ para MU y EX, respectivamente, son muy parecidos. Excepto en el caso del Pb para el origen EX, la dispersión de los valores es muy similar para ambas DOP.

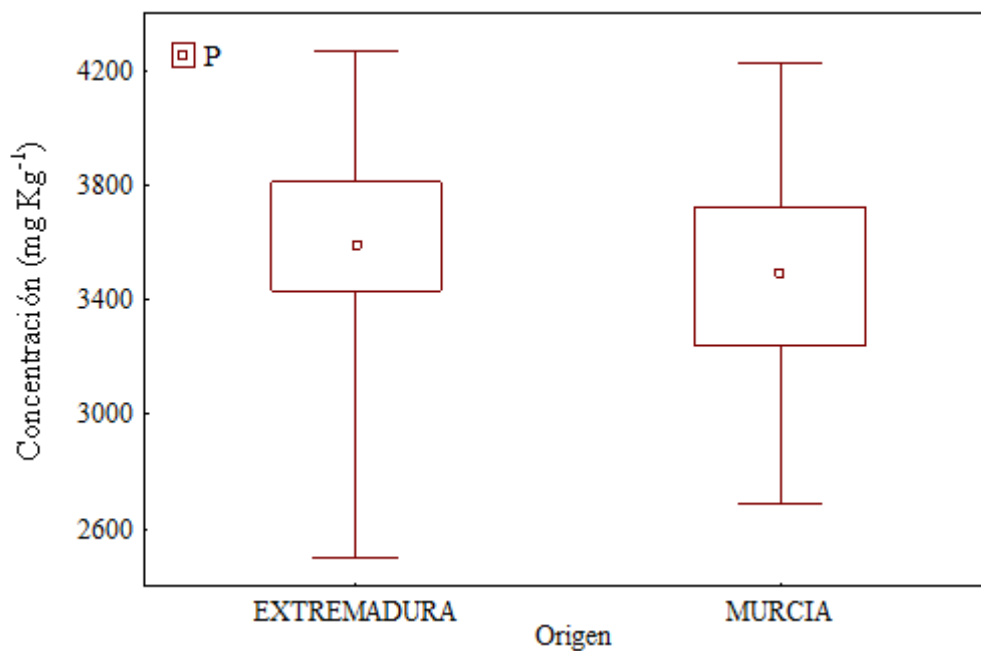


Figura IV.34. Diagrama de caja y bigotes para P.

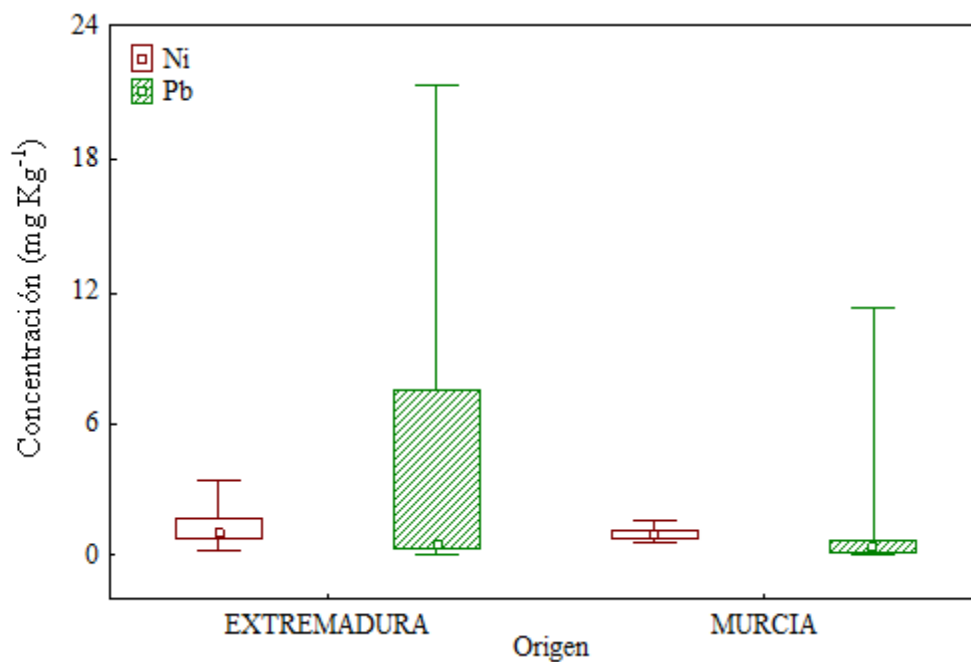


Figura IV.35. Diagrama de caja y bigotes para Pb y Ni.

Las diferencias observadas entre los distintos orígenes deben ser comprobadas estadísticamente mediante las correspondientes pruebas de significación. La secuencia operativa para determinar la normalidad o no de las muestras y las diferencias entre tipos se recogen en el esquema de la figura IV.36. Primero se realiza una prueba de normalidad para ver qué tipo de estadística debe utilizarse, paramétrica o no paramétrica. Para ello se lleva a cabo la prueba de Lilliefors [17]. El siguiente paso es comprobar si se pueden analizar los tres tipos de pimentón en conjunto o si existen diferencias entre los pimentones dulces, picantes y agridulces. Para ello, se aplica un ANOVA en el caso que la prueba de Lilliefors concluyera con que la estadística a realizar es paramétrica, o un test de Kruskal- Wallis en caso contrario. Si resulta que no existen diferencias entre los tipos, la diferenciación geográfica MU-EX puede hacerse sin distinguir tipos. No obstante, en caso contrario, la diferenciación geográfica MU-EX debe hacerse distinguiendo entre tipos, dulce, picante y agridulce.

Los resultados de la prueba de Lilliefors para comprobar si los datos se ajustan a una distribución normal se han incluido en la tabla IV.10, que muestra los valores calculados del parámetro estadístico L. Estos valores se comparan con los valores tabulados, y si el valor calculado es menor que el tabulado se puede considerar que los datos se ajustan a una distribución normal. En caso contrario habría que utilizar estadística no paramétrica. Según se aprecia en la tabla IV.10., los valores calculados del estadístico L son menores que el valor tabulado, L_{tab} para $\alpha=0,01$, para los elementos B, Ca, K, Mg, Na, P y Zn en el caso del tipo dulce. En el caso del tipo agridulce los elementos que presentan distribución normal son K y P. Para el pimentón del tipo picante los elementos que presentan valores de L_{cal} menor que L_{tab} , y por tanto distribución normal, son B, Mg, Na y P. La distribución no es normal para los demás los elementos. Por lo tanto, se considera que, en conjunto, los datos no tienen una distribución normal y se utilizará estadística no paramétrica para las posteriores comparaciones.

Tabla IV.10. Test de Lilliefors para normalidad.

Elemento	$L_{cal} D^a$	$L_{cal} A^a$	$L_{cal} P^b$
Al	0,13	0,19	0,21
B	0,12	0,16	0,10
Ca	0,09	0,14	0,17
Cu	0,21	0,26	0,19
Fe	0,19	0,19	0,22
K	0,12	0,12	0,15
Mg	0,09	0,14	0,06
Mn	0,24	0,19	0,14
Na	0,12	0,14	0,11
Ni	0,20	0,29	0,17
P	0,11	0,07	0,09
Pb	0,30	0,41	0,36
Sr	0,15	0,21	0,24
Zn	0,09	0,27	0,15

D: dulce; P: picante; A: agridulce. ^a $L_{tab}=0,13$; ^b $L_{tab}=0,12$

El estudio estadístico no paramétrico para la comparación entre los valores correspondientes a los distintos tipos se ha llevado a cabo mediante el test de Kruskal-Wallis [18]. En la tabla IV.11.se pueden observar los valores del estadístico H calculado, así como el análisis *post hoc*, marcando las comparaciones en las que hay diferencias significativas. Se han obtenido valores de H mayores que el valor tabulado, $\chi^2 = 5,99$, para los elementos Al, B, Ca, Fe, Mg, Mn, Ni, P y Pb. Los valores de H mayores se obtuvieron para Al, B, Fe y Mg, siendo estos elementos los que presentan diferencias significativas para, al menos, dos de las tres comparaciones consideradas. Los tipos picante y agridulce presentan diferencias para los elementos Al, B, Ca, Fe, Mg, Mn, Ni y Pb, mientras que los tipos agridulce y dulce presentan diferencias para Al, B, Fe, Mg y P. Ninguno de los elementos estudiados refleja diferencias entre los tipos picante y dulce. Los elementos Cu, K, Na, Sr y Zn presentan valores menores que χ^2 , para $\alpha=0,05$, y por tanto no hay diferencias significativas entre los resultados obtenidos para los mismos.

Tabla IV.11. Test de Kruskal-Wallis para la comprobación del contenido elemental en pimentones picantes, dulces y agridulces

Elemento	H	P-D	P-A	A-D
Al	15,11			
B	13,9			
Ca	12,18			
Cu	3,88			
Fe	21,98			
K	2,87			
Mg	15,35			
Mn	7,06			
Na	1,84			
Ni	13,54			
P	7,95			
Pb	10,39			
Sr	3,19			
Zn	0,81			

D: dulce; P: picante; A: agridulce. $\chi^2= 5,99$.

Puesto que el hecho de considerar en conjunto los tres tipos puede introducir una varianza importante a la hora de establecer los modelos de diferenciación geográfica, se procede a realizar el estudio de diferenciación del origen considerando los diferentes tipos. Además, con fines comparativos, se realizará una diferenciación global sin distinción de tipos.

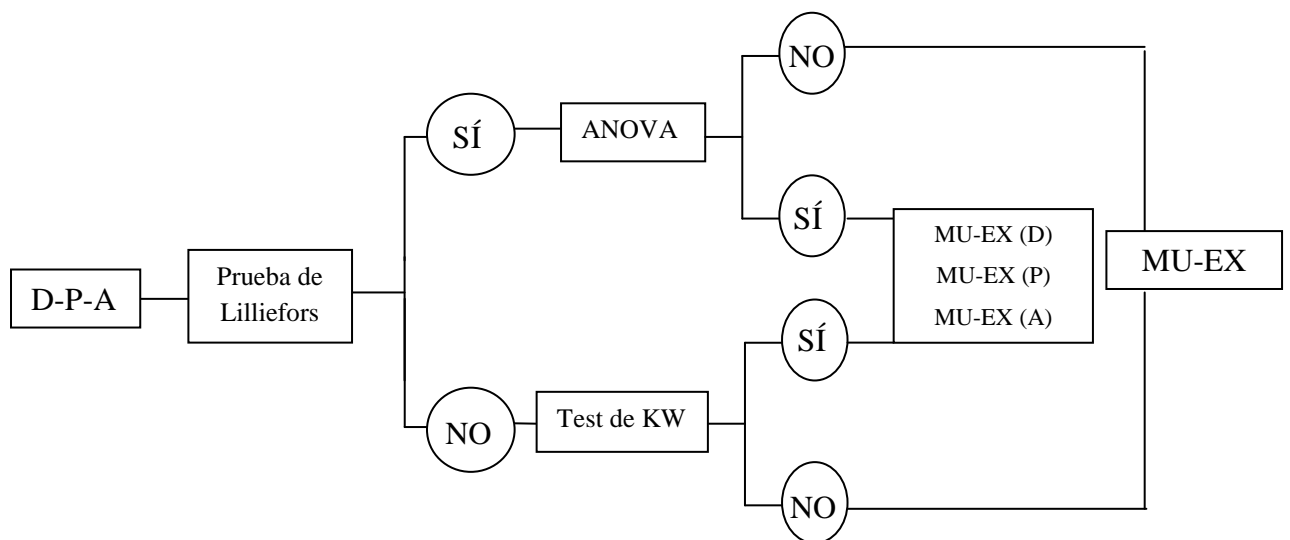


Figura IV.36. Secuencia operativa empleada para determinar si se discrimina entre tipos. D: dulce; P: picante; A: agridulce.

Antes de aplicar métodos de reconocimiento de patrones, se comprueba si los datos de los tipos dulce, picante y agridulce, dentro de cada origen, siguen una distribución normal. Para ello se aplica el test univariante de Lilliefors a cada uno de los elementos. Los valores de L_{cal} y L_{tab} se recogen en la tabla IV.12. Como la distribución no es normal para todos los elementos, se aplica una estadística no paramétrica.

Tabla IV.12. Test de Lilliefors para la normalidad.

Elemento	MURCIA			EXTREMADURA		
	$L_{cal} D^a$	$L_{cal} P^b$	$L_{cal} A^a$	$L_{cal} D^c$	$L_{cal} P^c$	$L_{cal} A^c$
Al	0,16	0,21	0,35	0,17	0,13	0,21
B	0,15	0,10	0,23	0,10	0,13	0,16
Ca	0,13	0,20	0,26	0,15	0,25	0,07
Cu	0,33	0,25	0,36	0,18	0,26	0,28
Fe	0,25	0,23	0,42	0,18	0,13	0,17
K	0,15	0,15	0,17	0,15	0,21	0,15
Mg	0,21	0,08	0,21	0,22	0,16	0,13
Mn	0,15	0,28	0,45	0,30	0,20	0,23
Na	0,19	0,12	0,24	0,10	0,16	0,13
Ni	0,21	0,22	0,21	0,20	0,19	0,23
P	0,21	0,15	0,13	0,15	0,09	0,12
Pb	0,33	0,45	0,46	0,30	0,28	0,41
Sr	0,19	0,31	0,26	0,29	0,25	0,23
Zn	0,19	0,12	0,37	0,18	0,22	0,19

^a $L_{tab}=0,13$; ^b $L_{tab}=0,12$; ^c $L_{tab}=0,16$.

Para ver qué elementos presentan poder discriminante dentro de los distintos orígenes, se realiza una prueba U de Mann-Whitney (tabla IV.13.), que es un ensayo similar a la t de Student, pero para datos que no tienen una distribución normal. La prueba U de Mann-Whitney nos revela qué elementos presentan mayor poder discriminante a la hora de diferenciar entre orígenes. En el caso de la diferenciación de los orígenes EX y MU para el tipo picante, encontramos que los elementos que presentan mayores valores de z son Al, Cu, Fe, Na, Pb y Sr. En el caso del tipo dulce, los elementos que mayor poder discriminante presentan son Cu, Fe, Mn, Na, P y Sr, y para el tipo agridulce, los valores mayores de z son para Al, Ca, Fe, Mg, Mn, Na, Ni, Pb y Sr. Si se realiza la prueba U de Mann-Whitney considerando los tres tipos en global, los elementos de mayor poder discriminante son Cu, Mg, Mn, Na, P, Pb y Sr.

Tabla IV.13. Prueba U de Mann-Whitney para los orígenes EX y MU

Elemento	Z _{cal}			
	P	D	A	G
Al	2,52	1,04	4,89	0,49
B	1,65	0,14	1,30	1,45
Ca	1,76	0,83	2,75	0,22
Cu	2,54	2,41	1,89	3,79
Fe	2,85	3,66	4,19	1,75
K	0,7	1,46	1,75	0,13
Mg	1,29	0,64	5,37	2,40
Mn	1,48	2,32	4,96	4,95
Na	2,58	4,82	3,09	6,07
Ni	0,38	0,16	4,14	1,47
P	0,05	3,96	0,36	2,34
Pb	2,42	0,36	3,39	2,51
Sr	2,91	4,23	3,23	5,94
Zn	0,1	1,02	0,20	1,00

D: dulce; P: picante; A: agridulce; G: global. $z_{tab}=1,96$

De acuerdo con estos resultados obtenidos se puede afirmar que existen diferencias significativas entre los valores de concentración obtenidos para las muestras pertenecientes a los dos orígenes geográficos considerados. Por lo tanto, resulta adecuada la aplicación de técnicas de reconocimiento de patrones para la obtención de modelos de clasificación que permitan la diferenciación de muestras provenientes de dichos orígenes.

Otra conclusión interesante de la aplicación de la prueba U de Mann-Whitney es que también puede ser utilizada como primera aproximación para excluir las variables con menos poder discriminante [19]. Por ello, en la sucesiva aplicación de técnicas de reconocimiento de patrones, se usarán como variables los elementos que tras aplicar la prueba U de Mann-Whitney presentan valores de z mayores de 1,96, que es el valor crítico para el parámetro z.

IV.2.2. Aplicación de técnicas de reconocimiento de patrones.

Una vez realizado el estudio estadístico básico, la secuencia operativa que se va a seguir para este estudio se describe en la figura IV.37. En primer lugar, se aplica como técnica de previsualización de la tendencia de los datos el análisis PCA en dos

modalidades. Por un lado se realiza un análisis PCA^a, donde las variables que se emplean son las que tras aplicar la prueba U de Mann Whitney se consideran más discriminantes. Por otro lado, se lleva a cabo un PCA^b usando todas las variables. El análisis PCA genera una serie de componentes principales (PC) a partir de las variables originales explicando la máxima varianza posible de la matriz de datos originales [20].

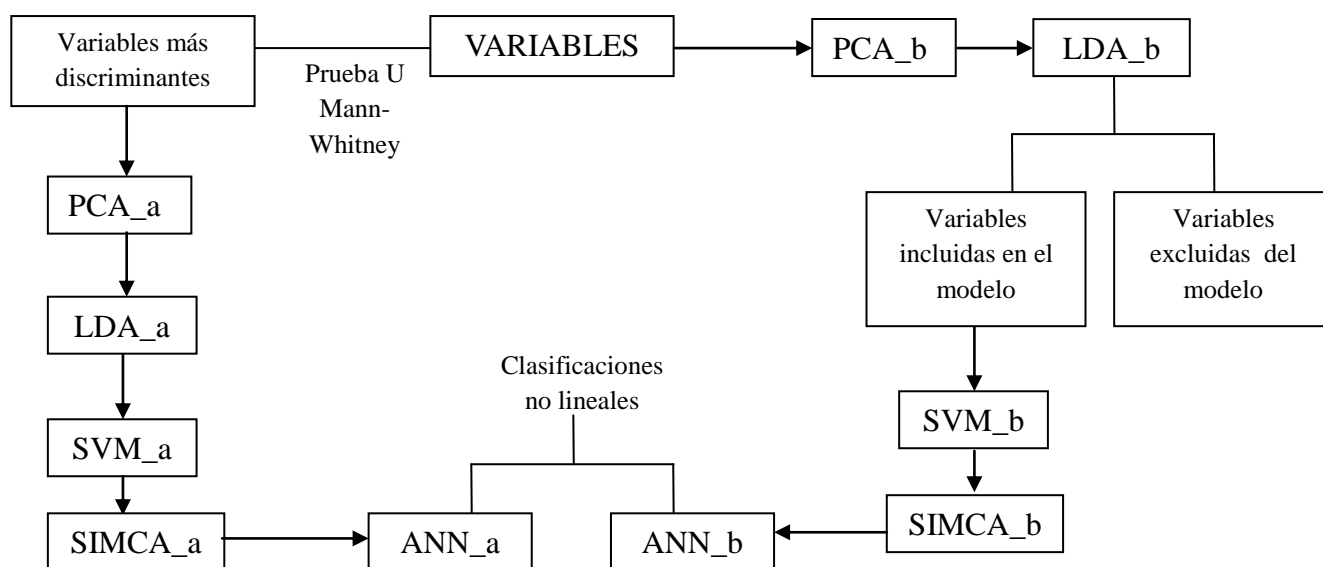


Figura IV.37. Secuencia operativa seguida en el análisis quimiométrico.

Posteriormente, se aplica análisis discriminante lineal (LDA), mediante el cual se calculan funciones discriminantes a partir de las combinaciones lineales de las variables originales, para diferenciar las clases consideradas minimizando la relación varianza dentro de la clase y la varianza entre clases [21]. El primer análisis LDA^a para establecer relaciones de pertenencia de los casos a las categorías estudiadas se construye con las variables iniciales más discriminantes según prueba U de Mann-Whitney. Por otro lado, se realiza un LDA^b, pero en este caso utilizando todas las variables obtenidas, para así seleccionar las variables con mayor poder discriminante, que se utilizarán en la aplicación de otras técnicas de clasificación.

Si la clasificación se puede realizar mediante modelos lineales se crean los modelos a partir de LDA, máquinas de vectores soporte (SVM) [22] lineales o modelado suave independiente de analogías de clases (SIMCA) [23]. Si por el contrario

la separación es compleja y no se puede realizar mediante un modelo lineal, hay que recurrir a técnicas como SVM no lineales o redes neuronales artificiales (ANN)[24].

Para hacer una validación cruzada de los modelos de clasificación se aplica el procedimiento de jacknife estratificado, eliminando un grupo de muestras (*Delete-A-Group-Jacknife, DAGJK*) [25], y creando nueve replicados para los conjuntos de entrenamiento (training) y prueba (test). DAGJK elimina de forma aleatoria un grupo de muestras de cada clase y construye el modelo. El grupo eliminado se usa como conjunto de prueba y se calcula la sensibilidad (SENS) y especificidad (SPEC) según las ecuaciones 4.25-26 [26].

$$SENS = \frac{TP}{TP + FN} \quad (4.25)$$

$$SPEC = \frac{TN}{TN + FP} \quad (4.26)$$

donde TN son los verdaderos negativos, TP los verdaderos positivos, FP los falsos positivos y FN los falsos negativos. En todos los casos se calcula la SENS y SPEC para los orígenes MU y EX, y una SENS y SPEC global ponderada teniendo en cuenta el número de muestras en cada origen.

IV.2.2.1. Diferenciación geográfica de pimentones dulces.

Se prepara una matriz de datos con 14 columnas, una por cada elemento determinado, y 45 filas, correspondientes a cada una de las muestras de pimentón dulce analizadas.

Se realizan dos análisis PCA, uno (PCA^a) utilizando como variables el contenido de los elementos Cu, Fe, Na, P, Mn y Sr, que son los señalados por la prueba U de Mann-Whitney, y otro (PCA^b) utilizando todas las variables. Después se realizará una comparación de los resultados obtenidos en ambos casos.

En el caso del PCA^a se obtienen los tres primeros componentes principales cuya expresión se incluye en las ecuaciones 4.27-29, que explican el 78,81% de la varianza de los datos originales. PC1 explica un 42,22%, PC2 un 22,09% y PC3 un 14,5 %. Los

dos primeros explican el 64,3% de la misma, y son los que se consideran puesto que sus autovalores son mayores que 1, 2,53 y 1,35 para PC1 y PC2, respectivamente, y según el criterio de Kaiser explican más varianza que una variable original [27].

$$PC1: -0,39|Fe|-0,31|P|+0,52|Cu|+0,17|Mn|+0,57|Sr|+0,36|Na|(4.27)$$

$$PC2: 0,33|Fe|-0,22|P|-0,32|Cu|-0,72|Mn|-0,11|Sr|+0,46|Na|(4.28)$$

$$PC3: 0,46|Fe|+0,86|P|-0,11|Cu|+0,04|Mn|+0,16|Sr|+0,11|Na|(4.29)$$

La distribución de las muestras de pimentón según su origen MU y EX en el espacio definido por los dos primeros PCs se puede ver en la figura IV.38. Si consideramos las coordenadas de PC1 y PC2, se observa que las muestras pertenecientes al origen EX están separadas de las del origen MU con valores negativos de PC2 y positivos y negativos de PC1. No se aprecia una separación completa de los dos orígenes, pero se puede observar que hay cierta tendencia a ello, con lo que cabe esperar que se pueda conseguir un modelo de clasificación adecuado para la diferenciación geográfica de ambas clases.

En cuanto a los pesos de las variables, hay que tener en cuenta que, si son superiores de 0,7, en valor absoluto, muestran una fuerte correlación; pesos entre 0,7 y 0,4 indican una participación moderada en la influencia de las variables a los PCs, y si tienen valores por debajo de 0,4 su influencia puede considerarse prácticamente nula [28]. La tabla IV.14. muestra los valores de los pesos de las variables obtenidos al aplicar PCA^a. En relación al PC1, las variables que tienen mayor poder discriminante son Cu, Sr y Fe, mientras que para el PC2 tenemos que las variables que mayor peso presentan son Mn, y Na. En el caso del PC3, el P es el que presenta mayor contribución.

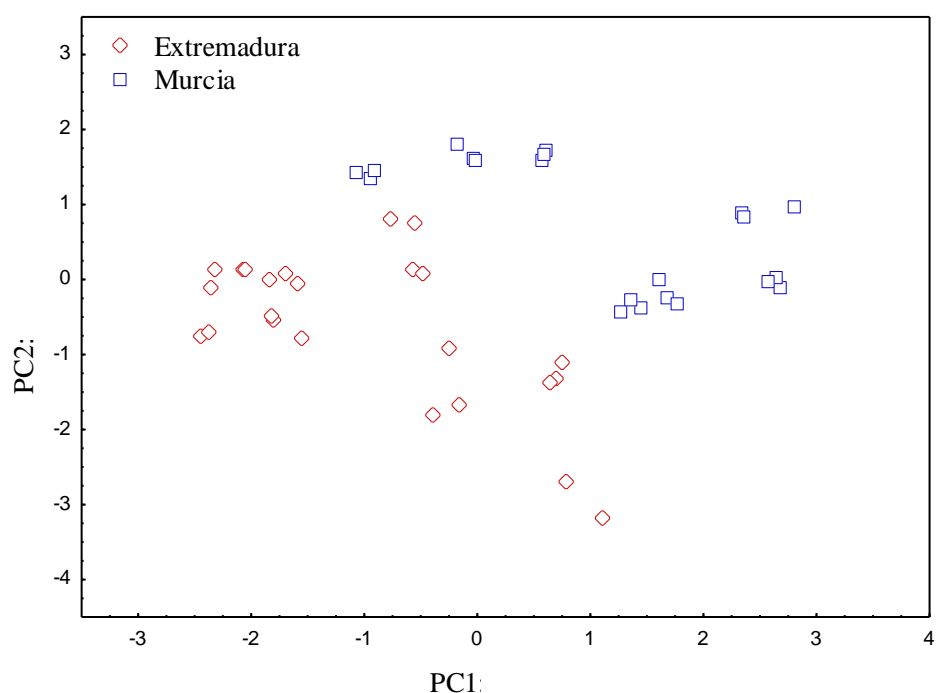


Figura IV.38. Distribución de muestras de pimentones dulces según su origen en el plano definido por los dos primeros PCs (modelo PCA^a)

Tabla IV.14. Peso de las variables en los PCs(modelo PCA^a).
Pimentón dulce

Elemento	PC1	PC2	PC3
Cu	0,83	-0,37	-0,10
Fe	0,62	0,38	0,42
Mn	0,28	-0,83	0,04
Na	0,57	0,53	0,10
P	-0,49	-0,25	0,79
Sr	0,91	-0,13	0,15

En el caso del PCA^b, considerando el contenido de todas las variables, se han calculado los tres primeros componentes principales, que explican el 64,67% de la varianza de los datos originales. PC1 explica un 39,66%, PC2 un 13,35% y PC3 un 11,66 %. Se consideran estos tres primeros PCs puesto que sus autovalores son mayores que 1, siendo 4,54, 2,37 y 2,05 para PC1, PC2 y PC3, respectivamente.

En este caso, las variables Al, B, Ca, Cu, Sr y Zn presentan un peso mayor en PC1, Fe y Zn están más correlacionados a PC2 y no existen correlaciones especialmente importantes de ningún elemento para PC3. En la tabla IV.15.se recogen estos pesos.

Tabla IV.15. Peso de las variables a los PC (modelo PC^b).

Pimentón dulce

Elemento	PC1	PC2	PC3
Al	-0,70	0,48	0,05
B	-0,70	-0,36	-0,04
Ca	-0,73	0,01	0,30
Cu	-0,86	-0,18	-0,25
Fe	-0,44	0,86	0,05
K	0,34	0,27	-0,56
Mg	-0,43	0,55	0,48
Mn	-0,65	-0,39	0,28
Na	-0,24	0,39	-0,64
Ni	0,39	0,13	-0,42
P	0,19	-0,06	0,58
Pb	0,26	0,32	0,58
Sr	-0,76	0,19	-0,01
Zn	-0,68	-0,62	-0,03

En la figura IV.39.se puede observar la distribución de las muestras en el espacio definido por los tres primeros PCs. La mayoría de las muestras del origen EX aparecen a valores negativos de PC2 y positivos de PC3. Las muestras del origen MU aparecen distribuidas para valores positivos y negativos de PC2 y PC3.Considerando la distribución con respecto al eje PC1, La mayoría de las muestras de EX se sitúan en la parte del gráfico que corresponde a los valores positivos de PC1, mientras que las muestras del origen MU se encuentran en la parte del gráfico correspondiente a valores negativos de PC1, aunque varias muestras también aparecen para valores positivos de PC1. Al igual que en cálculo realizado según PC^a, se aprecia una cierta separación de las muestras, aunque hay solapamiento de las mismas.

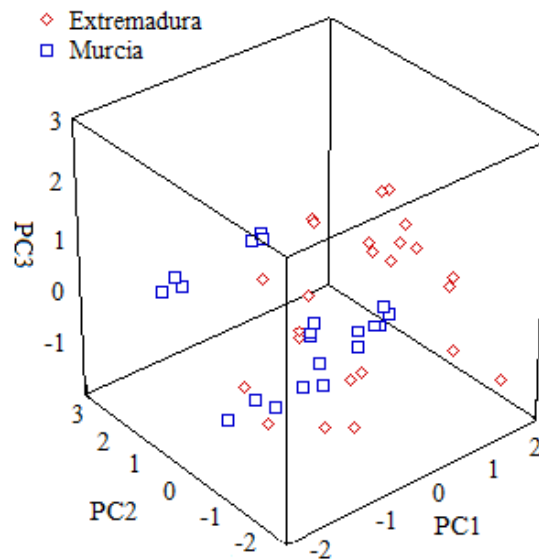


Figura IV.39. Distribución de muestras de pimentones dulces según su origen en el plano definido por los tres primeros PCs.

Una vez realizado el PCA se aplica una técnica de reconocimiento de patrones supervisado como el LDA, con el que pueden obtenerse reglas de clasificación adecuadas que permiten diferenciar las muestras. También es posible utilizar LDA como técnica exploratoria de datos para seleccionar las variables que más contribuyan al modelo de clasificación, y poder así eliminar las que menos contribuyan [29]. Nuevamente, para la realización del LDA es posible partir de las variables que tras aplicar la prueba U de Mann-Whitney presentaban diferencias significativas (LDA^a), o partir de todas las variables (LDA^b).

En el caso de LDA^a se aplicó un análisis sucesivo incluyendo progresivamente las variables más discriminantes para obtener las que más contribuyen a la diferenciación. Estas variables son: Mn, Sr, Na, Cu y P. En la figura IV.40. se observa la separación que se consigue entre los orígenes geográficos de las muestras utilizando esas variables. Como se puede apreciar hay una buena separación de las muestras correspondientes a las DOP EX y MU. La ecuación. 4.30 corresponde a la función discriminante (DF1), en la que se puede apreciar la contribución de cada variable.

$$DF1: 0,79|Na|+1,52|Sr|-1,04|Mn|-0,63|P|-0,60|Cu| \quad (4.30)$$

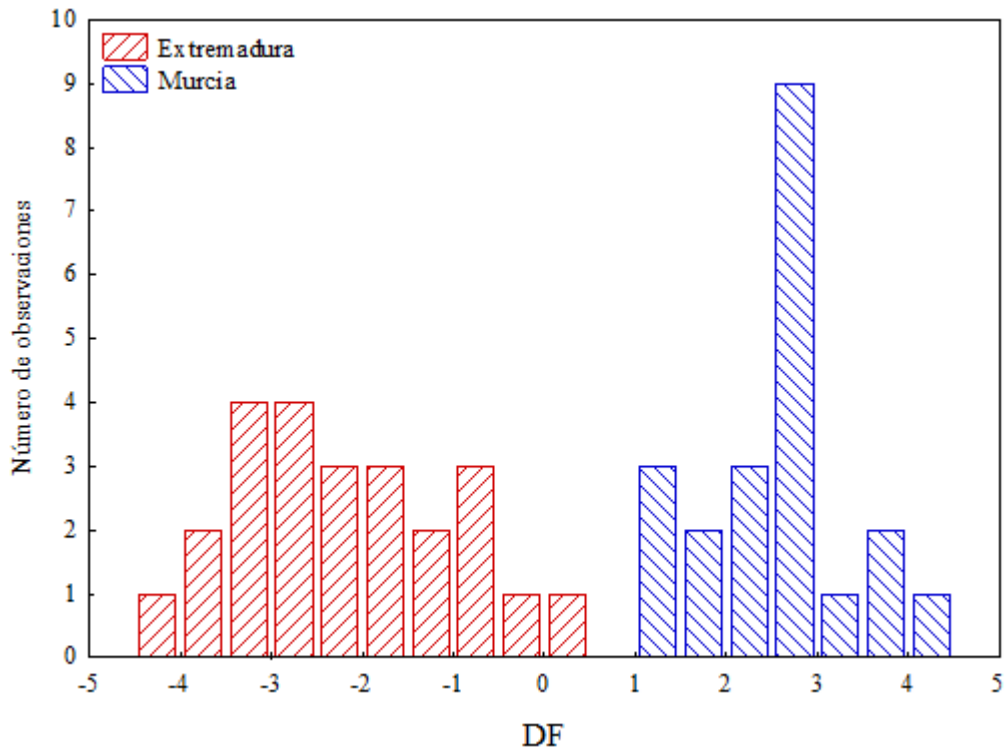


Figura IV.40. Distribución de las muestras de pimentones dulces en DF según el origen, aplicando modelo LDA^a.

La validación del modelo se realizó según el procedimiento DAGJK. En la tabla IV.16 se muestran los resultados globales de la capacidad del modelo creado para diferenciar orígenes de muestras de pimentón junto con los valores de SENS y SPEC. Los resultados globales se corresponden con una media ponderada de los resultados de SENS y SPEC de los orígenes geográficos MU y EX.

Como puede apreciarse, los resultados del modelo de clasificación obtenidos mediante LDA^a son buenos, ya que SENS y SPEC del conjunto de todas las muestras son 96,2 y 96,9, respectivamente.

Como ya se ha indicado, el cálculo de LDA^b se realizó utilizando de partida todas las variables. El procedimiento es similar al descrito anteriormente, ya que se aplicó un análisis sucesivo incluyendo progresivamente las variables más discriminantes, y para hacer la validación cruzada del modelo de clasificación se aplica igualmente el procedimiento DAGJK con nueve replicados. Las variables que se han eliminado en este caso son, Sr y Zn.

En la ecuación 4.31 se presenta la función discriminante (DF1), con la contribución de cada variable.

$$DF1: -2,22|Na|+1,98|Ni|-1,87|Fe|+1,35|P|+1,99|Mn|-1,75|Ca|-0,52|B|-1,15|Cu|+1,05|Pb|+0,86|Al|-0,39|K|+0,66|Mg| \quad (4.31)$$

En la figura IV.41. se representa la distribución de las variables según la función discriminante calculada. Como se puede apreciar hay una buena separación utilizando esta función discriminante con las variables indicadas. Se han calculado SENS y SPEC de los orígenes MU y EX, y los valores obtenidos se incluyen en la tabla IV.16.

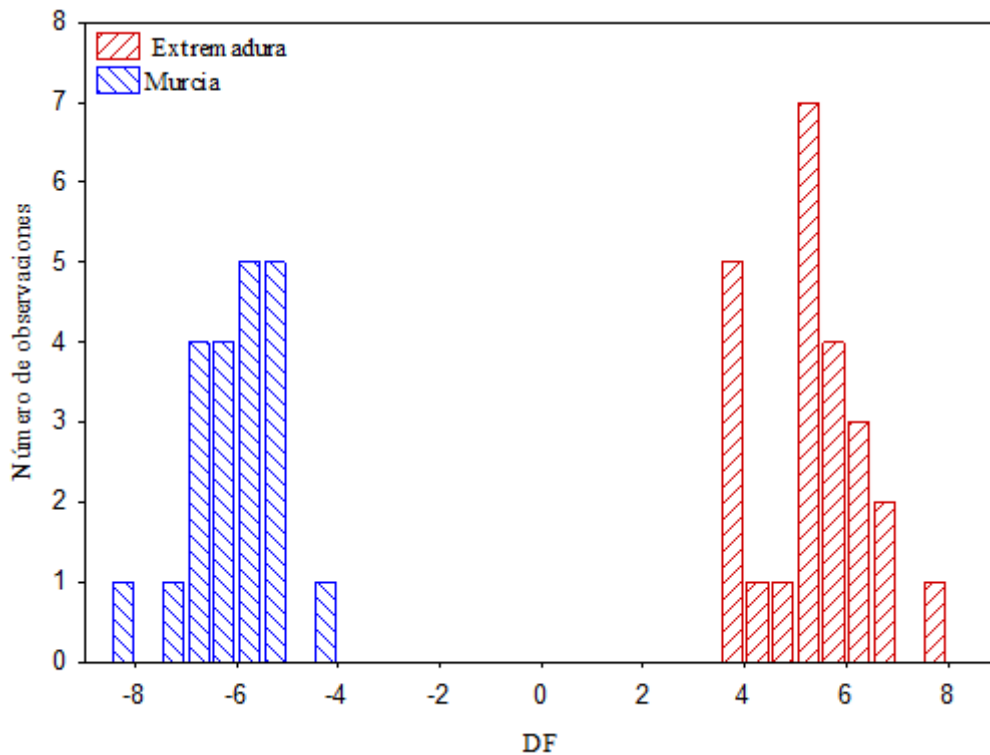


Figura IV.41. Distribución de las muestras de pimentones dulces en DF según el origen, aplicando modelo LDA^b.

Siguiendo con el esquema propuesto, y para intentar conseguir modelos más eficaces, se ha aplicado la técnica SVM lineal. En este caso también se realizan dos modelos, a y b, con las mismas variables originales que los modelos de LDA descritos.

Al igual que para el LDA, el modelo de clasificación debe ser construido a partir de un conjunto de muestras de entrenamiento y otro de prueba. De manera análoga se

dividió el conjunto de muestras en un 70% para muestras de entrenamiento y un 30% para muestras de prueba tanto para SVM^a como SVM^b. El modelo se sometió a la validación cruzada aplicando la prueba DAGJK. Los resultados globales de predicción de los modelos SVM se muestran en la tabla IV.16.

Como puede observarse en la tabla IV.16., el modelo obtenido con SVM^a mejora ligeramente los resultados obtenidos mediante LDA^a. SENS y SPEC son 98% y 97%, respectivamente. En el caso del modelo SVM^b, la SENS obtenida para las muestras de pimentón dulce de Murcia es menor que cuando el modelo se obtiene mediante LDA^b.

Tabla IV.16. Resultados de los modelos de clasificación global para muestras de pimentón dulce

	MU		EX		Global (MU-EX)	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA ^a	93 ± 11	100	100	93 ± 11	96 ± 6	97 ± 5
LDA ^b	100	100	100	100	100	100
SVM ^a	97 ± 10	97 ± 6	97 ± 6	97 ± 10	98 ± 6	97 ± 5
SVM ^b	100	98 ± 7	98 ± 7	100	99 ± 3	99 ± 4
SIMCA ^a	84 ± 9	96 ± 7	97 ± 7	100	91 ± 3	99 ± 4
SIMCA ^b	89 ± 15	100	76 ± 15	100	83 ± 10	100
ANN ^a	100	100	100	100	100	100
ANN ^b	100	100	100	100	100	100

MU: Murcia; EX: Extremadura; G: Global. SENS: sensibilidad; SPEC: especificidad. a: Mann-Whitney. b: LDA con todas las variables y SVM, SIMCA y ANN con las variables incluidas al aplicar el modelo por LDA^b.

SIMCA realiza un análisis PCA para cada clase considerada determinando el número de PCs necesarios para describir la estructura de cada clase. SIMCA permite una visualización muy eficaz de la clasificación de las muestras, así como de los falsos positivos y negativos, ya que al ser una técnica de modelado suave permite que una muestra pueda pertenecer o no a una clase, y también clasificarse en el espacio confluyente entre las clases. En este caso, la sensibilidad es la proporción de muestras que pertenecen a una y se clasifican correctamente. La especificidad es la proporción de muestras que no pertenecen a una clase determinada y son identificadas como extrañas por el modelo [30]. La SPEC aumenta cuando muestras de una clase que no se asignan dentro de su clase, no son asignadas a otras distintas, sino como pertenecientes a ninguna clase.

Tal y como se ha hecho hasta ahora, se calcularon dos modelos de clasificación con SIMCA. En uno donde se eligen las variables consideradas como más discriminantes según la prueba Mann-Whitney, SIMCA^a. Se calcula otro modelo SIMCA empleando las variables que se utilizaron para calcular el modelo LDA^b, que denominamos como SIMCA^b.

En el caso de SIMCA^a las variables que se emplearon fueron Cu, Fe, Mn, Na, P y Sr. En la tabla IV.16, donde se muestran los resultados, se puede observar cómo SPEC mejora cuando se trata de los resultados del origen EX y también para los globales. En este caso de la SENS disminuye porque existen muestras tanto de EX como de MU que no se asignan a ninguna clase.

En el gráfico de la figura IV.42., denominado gráfico de Cooman, se representan las distancias críticas que definen un modelo y otro y las distancias de las muestras a esos dos modelos.

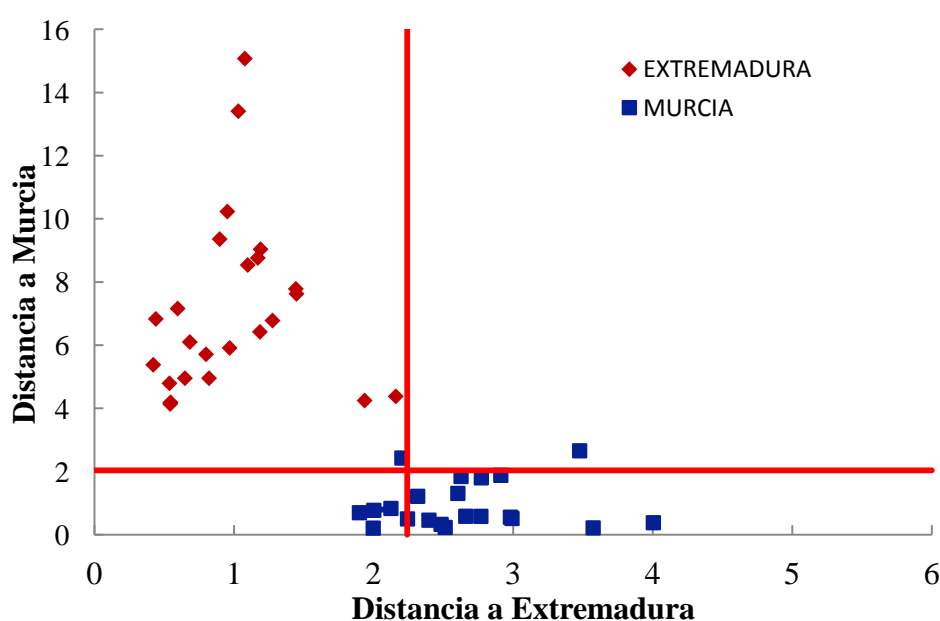


Figura IV.42.Gráfico de Cooman pimentón tipo dulce según modelo SIMCA^a.

Para el cálculo del modelo SIMCA^b se tomaron las variables elegidas mediante LDA^b, es decir, todas las variables iniciales a excepción del Sr y Zn. La figura IV.43 muestra el gráfico de Cooman. Se aprecia que todas las muestras del origen EX están clasificadas como EX, mientras que sólo una del origen MU está clasificada como no

clase, estando todas las demás dentro del origen MU. Según esta apreciación SPEC y SENS para EX sería 100%. No obstante, al realizar la validación cruzada con distintos grupos de muestras no se obtiene una eficacia tan elevada. Los valores de SPEC y SENS para ambos orígenes se encuentran en la tabla IV.16. Para el origen EX son 100 y 76%, respectivamente, y para el origen MU son 100 y 89%.

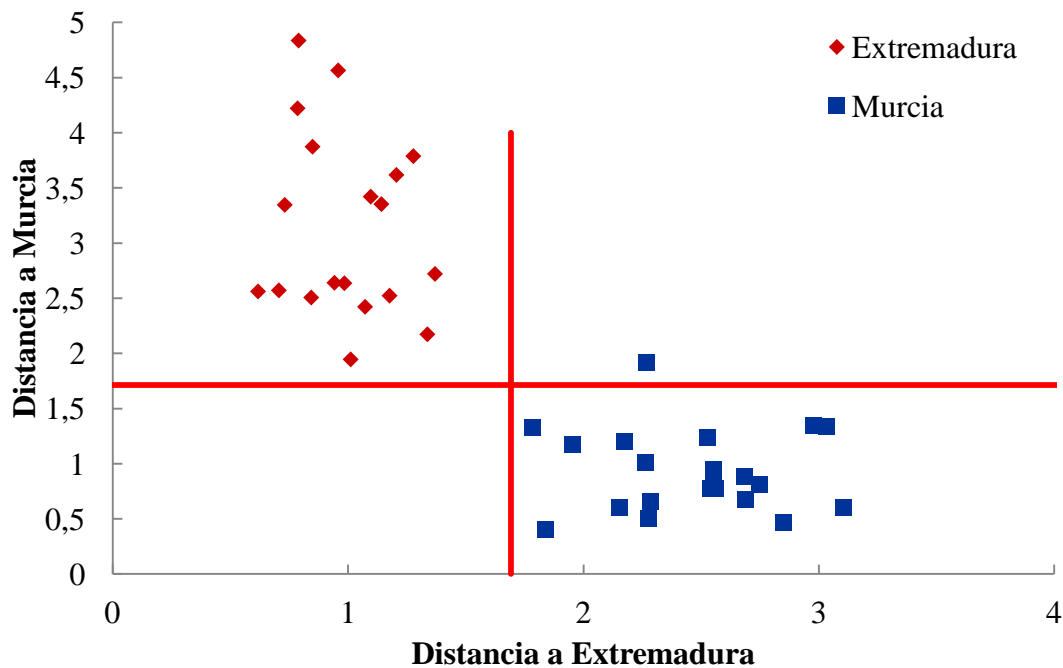


Figura IV.43. Gráfico de Cooman pimentón dulce según modelo SIMCA^b.

Una vez aplicados métodos lineales como LDA, SVM y SIMCA para obtener modelos de clasificación de las muestras de pimentón dulce, y con fines comparativos se han aplicado métodos no lineales como son las redes neuronales artificiales, ANN. Las redes que se emplean son redes de perceptrones multicapas (MLP) que consisten en neuronas organizadas en capas, una de entrada correspondiente a las variables elegidas para crear el modelo, otra oculta donde cada neurona realiza una suma ponderada de las variables de entrada y las transforma en una respuesta de salida, y otra capa de salida donde se calcula la probabilidad de pertenecer a una clase u otra. Las MLP-ANN se entrenan mediante un algoritmo de retropropagación, por lo que es necesario un

conjunto de muestras de verificación. Las muestras se dividen en un 50%, 25% y 25% para entrenamiento, verificación y test, respectivamente.

Se realizan dos modelos MLP-ANN, uno empleando las variables que permanecen tras aplicar el test de Mann-Whitney (MLP-ANN^a), y otro con las variables que LDA^b detectó como más discriminantes (MLP-ANN^b). La arquitectura de la MLP-ANN^a es 6:7:2 y la de la MLP-ANN^b 12:7:2. En la tabla IV.16. se muestran los resultados obtenidos Ambos tipos de redes permiten clasificar las muestras de pimentón dulce, obteniéndose resultados del 100% con independencia de las variables empleadas en la construcción del modelo.

IV.2.2.2. Diferenciación geográfica de pimentones picantes.

El primer paso es preparar una matriz de datos con 14 columnas, una por cada elemento analizado, y 64 filas correspondientes a cada una de las muestras de pimentón picante estudiadas.

Siguiendo el esquema de la figura IV.37 se realiza, en primer lugar, un análisis PCA^a, utilizando como variables el contenido de los elementos que según la prueba U de Mann-Whitney son buenos descriptores, ya que presentan diferencias significativas. Por otro lado, se realiza un análisis PCA^b empleando todas las variables.

En el caso de PCA^a, los elementos que poseen más poder discriminante son Al, Cu, Fe, Na, Pb y Sr. El 89,7% de la varianza de los datos originales se puede expresar con los tres primeros PCs, cuyas expresiones se incluyen en las ecuaciones. 4.32-4.34. Los dos primeros explican el 77,3% de la misma, el PC1 un 54,8% y el PC2 un 20,5%. PC3 explica un 14,4 % de la varianza. En este caso se consideran los dos primeros PCs puesto que sus autovalores son mayores que 1, siendo 3,29 y 1,23 para PC1y PC2 respectivamente.

$$PC1:0,50|Al|-0,51|Fe|-0,48|Cu|-0,24|Pb|-0,45|Sr|-0,13|Na|(4.32)$$

$$PC2:-0,12|Al|-0,03|Fe|+0,29|Cu|-0,26|Pb|-0,036|Sr|+0,83|Na|(4.33)$$

$$PC3:0,29|Al|+0,18|Fe|-0,02|Cu|-0,91|Pb|+0,02|Sr|-0,23|Na|(4.34)$$

En la figura IV.44.se representa la distribución de las muestras de pimentón dulce según su origen MU y EX en el espacio definido por los dos primeros PCs. Las muestras de MU se distribuyen, en su mayoría, en la zona de valores positivos de PC2 y negativos y positivos de PC1. Las muestras de EX se distribuyen con valores negativos de PC1 y PC2. En la zona de valores negativos de PC2 y positivos de PC1 se mezclan algunas muestras de EX y MU. No se aprecia una separación completa de los dos orígenes, pero sí es posible advertir cierta tendencia a la separación de las muestras de los distintos orígenes.

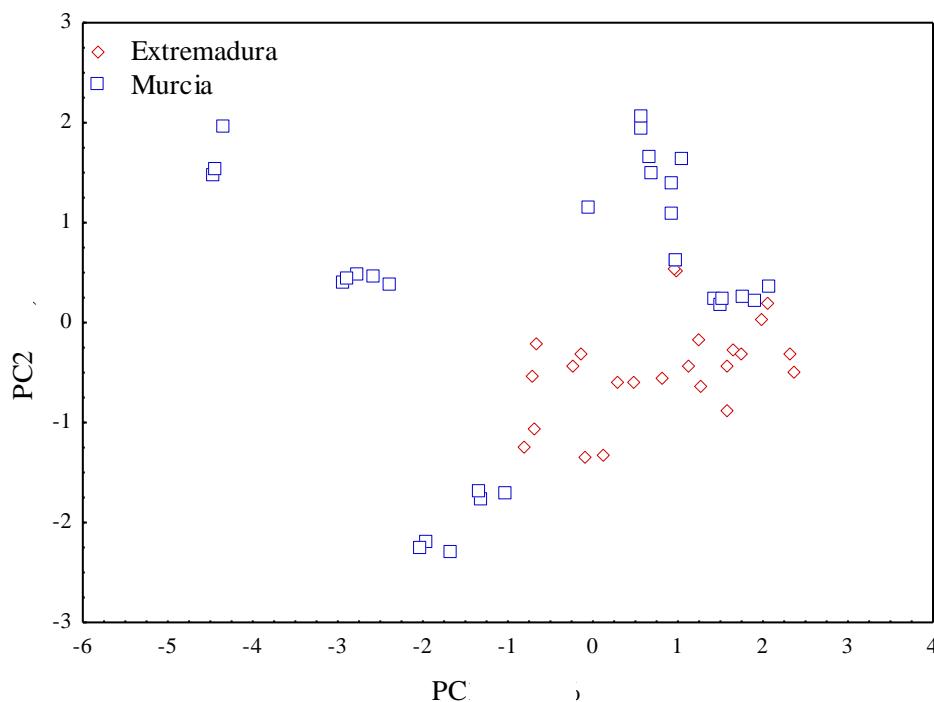


Figura IV.44. Distribución de muestras de pimentón picante según su origen en el plano definido por los dos primeros PCs.

Atendiendo al peso de las variables originales en el primer componente principal, las que mayor poder discriminante tienen son Al, Cu, Fe, y Sr, mientras que si observamos PC2 tenemos que las variables que mayor contribución presentan son Sr, y Na. En el caso del tercer PC la variable que más peso tiene es Pb, lo cual se puede observar en la tabla IV.17.

Tabla IV.17. Peso de las variables en los PCs (modelo PCA^a).
Pimentón picante

Elemento	PC1	PC2	PC3
Al	-0,89	-0,14	0,27
Fe	-0,93	0,00	0,17
Cu	-0,86	-0,33	0,02
Sr	-0,81	0,40	0,00
Pb	-0,43	-0,29	-0,85
Na	-0,23	0,93	-0,21

En el caso del PCA^b, los tres primeros PCs explican el 71,3% de la varianza total. Las expresiones de cada uno de los tres PCs se recogen en las ecuaciones 4.35-4.37. Se consideran estos tres PCs ya que sus autovalores son mayores que 1, con los valores 6,23, 2,27 y 1,49 para PC1, PC2 y PC3, respectivamente.

$$PC1: -0,32|Al| - 0,24|B| - 0,30|Ca| - 0,38|Cu| - 0,31|Fe| + 0,17|K| - 0,24|Mg| - 0,35|Mn| + 0,05|Na| - 0,22|Ni| + 0,13|P| - 0,20|Pb| - 0,26|Sr| - 0,35|Zn| \quad (4.35)$$

$$PC2: -0,22|Al| + 0,16|B| - 0,10|Ca| + 0,02|Cu| - 0,31|Fe| - 0,28|K| - 0,30|Mg| + 0,26|Mn| - 0,51|Na| + 0,22|Ni| - 0,22|P| + 0,12|Pb| - 0,34|Sr| + 0,18|Zn| \quad (4.36)$$

$$PC3: 0,14|Al| - 0,45|B| - 0,12|Ca| - 0,04|Cu| + 0,15|Fe| + 0,19|K| - 0,12|Mg| + 0,03|Mn| + 0,15|Na| + 0,12|Ni| - 0,63|P| + 0,47|Pb| - 0,06|Sr| - 0,17|Zn| \quad (4.37)$$

Los pesos de las variables en los tres PCs considerados se muestran en la tabla IV.18. Los elementos que más contribuyen a PC1 son Al, Ca, Cu, Fe, Mn y Zn, ya que presentan pesos superiores a 0,7. En el segundo componente principal, el elemento Na tiene un peso superior a 0,7, mientras que en el tercer componente principal es el P. Por lo tanto, para este modelo, éstas son las variables que más discriminan ya que son las que presentan los pesos mayores en los tres primeros componentes principales. Para visualizar las tendencias definidas por el análisis en componentes principales PCA^b se representa la distribución de las muestras de pimentón picante en el espacio de los tres primeros PCs.

Tabla IV.18. Peso de las variables en los PCs. (modelo PCA^b).
Pimentón picante

Elemento	PC1	PC2	PC3
Al	-0,80	-0,33	0,17
B	-0,61	0,23	-0,55
Ca	-0,75	-0,15	-0,15
Cu	-0,95	0,03	-0,04
Fe	-0,78	-0,46	0,19
K	0,43	-0,41	0,23
Mg	-0,59	-0,58	-0,14
Mn	-0,87	0,40	0,03
Na	0,12	-0,77	0,18
Ni	-0,55	0,34	0,15
P	0,33	-0,33	-0,77
Pb	-0,50	0,18	0,57
Sr	-0,64	-0,51	-0,08
Zn	-0,87	0,28	-0,21

En la figura IV.45. se muestra esta distribución, en la que se no aprecia una separación de las muestras de origen diferente, pero hay ciertas tendencias. Las muestras de EX están distribuidas principalmente en la zona positiva de PC2, mientras que las de MU lo hacen a valores negativos de PC2.

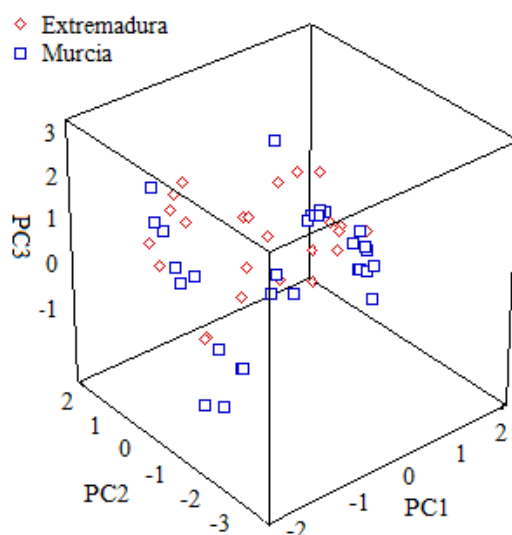


Figura IV.45. Distribución de muestras de pimentón picante según su origen en el plano definido por los tres primeros PCs.

Una vez realizado el PCA se aplica el LDA, utilizando, en primer lugar, las variables que tras aplicar la prueba U de Mann-Whitney presentaban diferencias significativas, LDA^a. Por otra parte, y siguiendo un esquema de trabajo similar al utilizado para las muestras de pimentón dulce, se llevará a cabo un LDA con todas las variables, LDA^b. En ambos casos se ha aplicado un análisis sucesivo incluyendo progresivamente las variables más discriminantes para obtener las que más contribuyan a la diferenciación. En el caso de LDA^a, las variables más discriminantes son Fe, Na y Pb. En la figura IV.46. se muestra la distribución de las muestras de pimentón picante según la función discriminante calculada. Como puede verse, no hay una diferenciación total, ya que existe un solapamiento de ambos orígenes, lo que indica que son necesarias otras técnicas para la correcta clasificación. En la ecuación 4.38 se presenta la función discriminante con la contribución de cada variable.

$$DF1: -0,48|Na|-0,95|Fe|+0,85|Pb| \quad (4.38)$$

En la tabla IV.19. se muestran los resultados de la capacidad del modelo creado para diferenciar orígenes de muestras de pimentón picante junto con la SENS y SPEC del modelo.

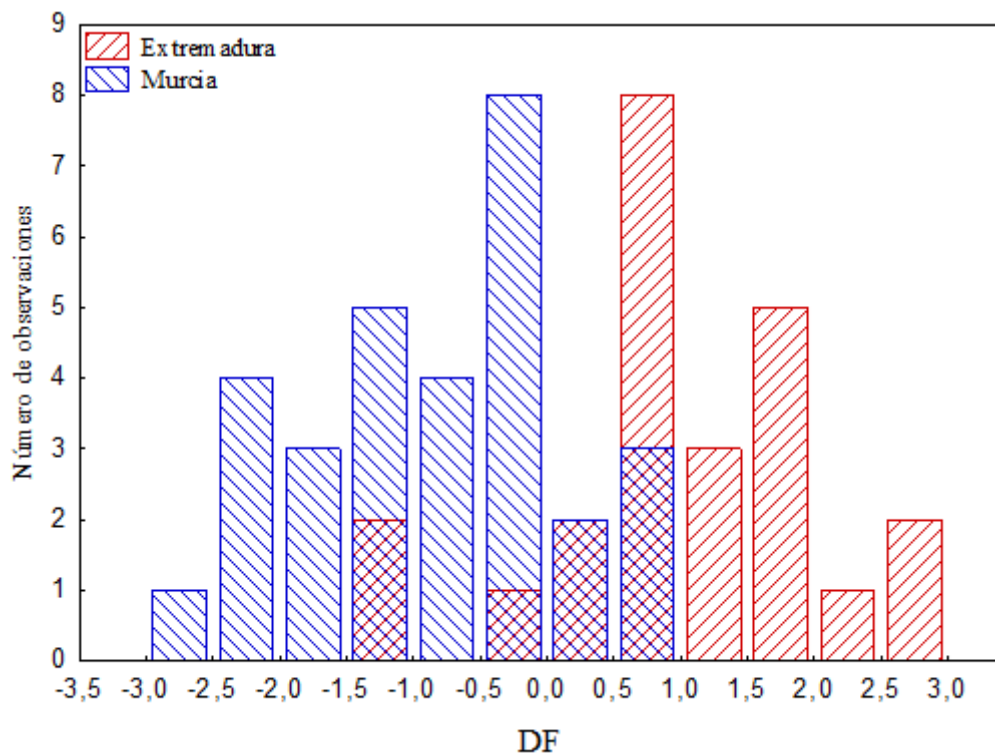


Figura IV.46. Distribución de las muestras de pimentones picantes en DF según el origen, aplicando modelo LDA^a

Tabla IV.19. Resultados de los modelos de clasificación global para muestras de pimentones picantes.

	MU		EX		Global (MU-EX)	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA ^a	74 ± 14	84 ± 8	84 ± 8	74 ± 14	80 ± 9	78 ± 10
LDA ^b	89±13	93±12	93±12	89±13	91±10	91±10
SVM ^a	72 ± 19	89 ± 7	89 ± 7	72 ± 19	82 ± 6	79 ± 10
SVM ^b	87±18	83±12	83±12	87±18	85±9	85±8
SIMCA ^a	83 ± 25	26 ± 19	76 ± 24	81 ± 34	80 ± 14	55 ± 20
SIMCA ^b	89±16	35±21	78±20	88±20	85±11	59±19
MLP-ANN-6:7:2 ^a	97 ± 6	98 ± 5	98 ± 5	97 ± 6	97 ± 4	98 ± 3
MLP-ANN-10:7:2 ^b	99 ± 4	96 ± 7	96 ± 7	99 ± 4	98 ± 4	97 ± 4

MU: Murcia; EX: Extremadura; G: Global. a: Mann-Whitney. b: LDA con todas las variables y SVM, SIMCA y ANN con las variables incluidas al aplicar el modelo por LDA^b.

Si analizamos los datos para LDA^b, encontramos que las variables que se eliminan del modelo son Al, Na, Ni y P. La ecuación de la función discriminante es la que se muestra a continuación.

$$DF1: -0,87|Fe| + 0,69|Pb| + 0,18|B| - 0,25|Sr| - 2,25|Cu| + 1|Mn| + 1,29|Zn| + 0,97|Mg| - 0,52|Ca| - 0,38|K| \quad (4.39)$$

En la figura IV.47. se muestra la distribución de las muestras de pimentón picante pertenecientes a las dos denominaciones de origen consideradas según la función discriminante calculada. Las muestras del origen MU aparecen en la parte izquierda del gráfico, que corresponde a valores negativos de la función discriminante, mientras que las del origen EX se sitúan en la parte derecha, que corresponde a valores positivos de la función discriminante. Puede apreciarse un solapamiento en la zona central, lo cual indica que la diferenciación no es adecuada. Este comportamiento es similar al que se obtuvo con el modelo LDA^a para pimentones picantes.

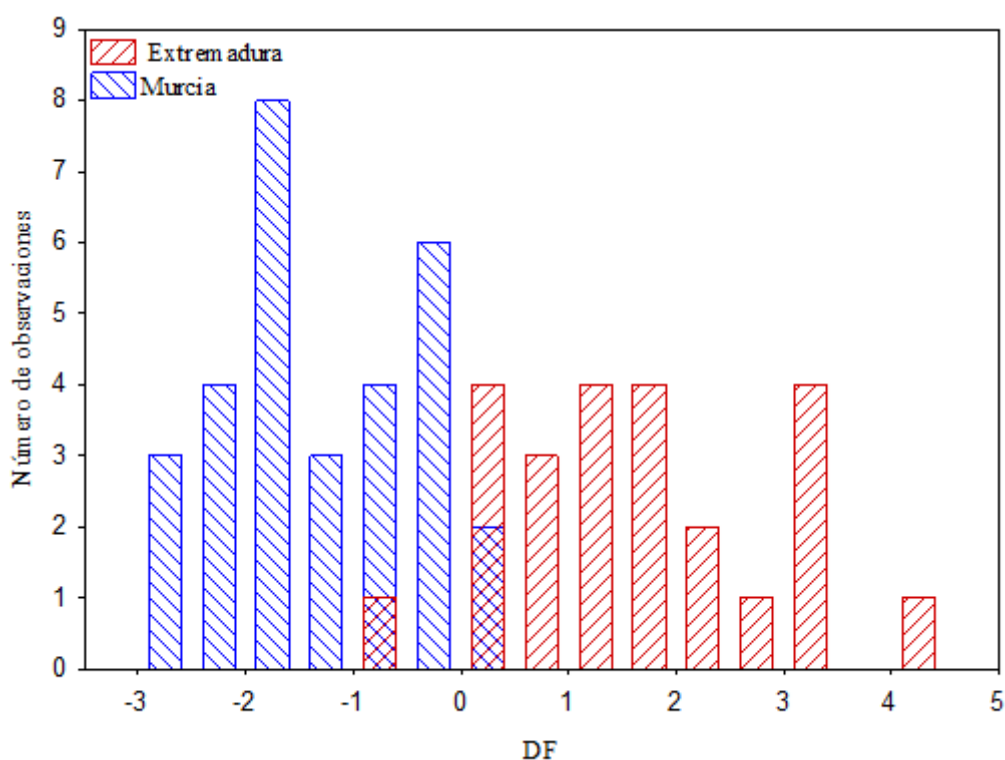


Figura IV.47. Distribución de las muestras de pimentones picantes en DF según el origen, aplicando modelo LDA^b.

Se realiza la validación del modelo siguiendo el procedimiento DAGJK. Los resultados obtenidos para los parámetros SENS y SPEC se han incluido en la tabla IV.19. El modelo LDA^b da valores mayores, que son iguales al 91% cuando se tiene en cuenta el conjunto de las dos denominaciones de origen.

Los resultados de SENS y SPEC del modelo LDA^b son mejores que los correspondientes al modelo LDA^a. No obstante, para intentar conseguir valores de SENS y SPEC más próximos al 100%, y con fines comparativos se han aplicado otros procedimientos de cálculo basados en las técnicas SVM, SIMCA y ANN.

Se crea un modelo a partir de SVM lineal de acuerdo a la prueba U de Mann-Whitney (SVM^a) y otro a partir de las variables que LDA^b determinó como más discriminantes (SVM^b). Al igual que en el caso de ambos LDA, el modelo de clasificación debe ser construido a partir de un conjunto de muestras de entrenamiento (70%) y otro de prueba (30%). Los resultados de SENS y SPEC se muestran en la tabla IV.19. Utilizando las SVM no se consiguen unos resultados que mejoren

sustancialmente los de LDA, por lo que se recurre a otra técnica para crear otro modelo de clasificación.

Los resultados del modelo SIMCA^a se muestran en la tabla IV.19. Tanto en la tabla de resultados como en el gráfico de Cooman (figura IV.48.), se observa que los resultados no responden a una buena clasificación. En el gráfico se aprecia cómo muestras de MU y EX son asignadas a ambos orígenes y muestras de EX se asignan como MU. En el caso de SIMCA^b, donde se utilizan para construir el modelo todas las variables originales excepto Al, Na, Ni y P, tampoco se obtienen buenos resultados. El gráfico de Cooman de la figura IV.49. muestra claramente cómo la separación no es buena, ya que hay muestras asignadas a ambos orígenes y muestras asignadas en orígenes cambiados.

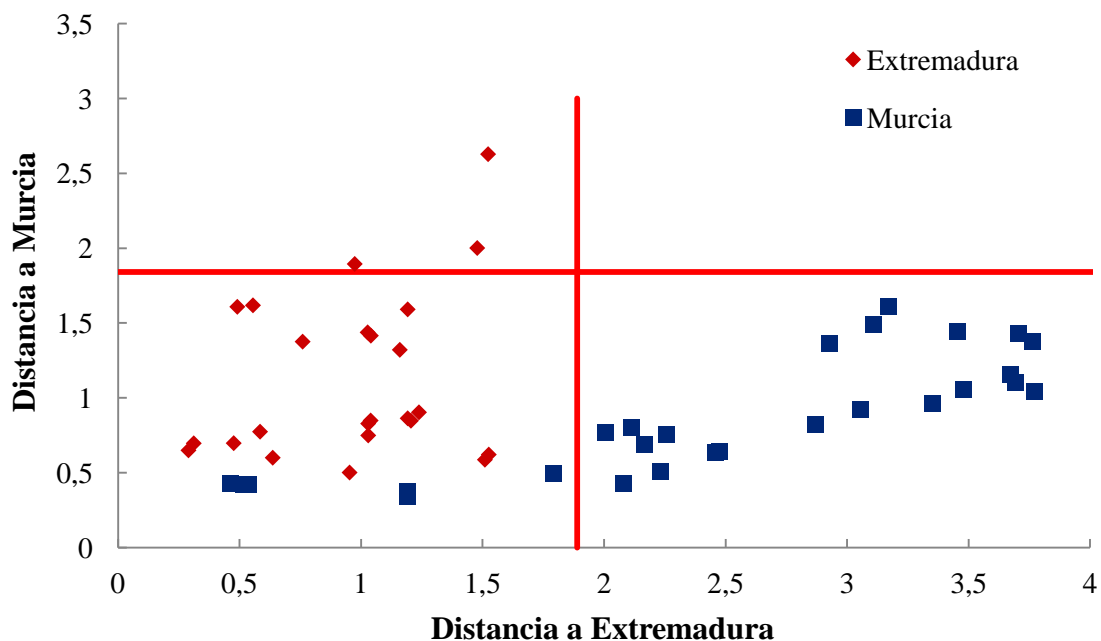


Figura IV.48. Gráfico de Cooman para las muestras del tipo picante según el modelo SIMCA^a.

En la tabla IV.19. se presentan los valores de SENS y SPEC encontrando que no es posible una separación mediante técnicas de reconocimiento de patrones lineales, y por ello se recurre a modelos no lineales como MLP-ANN.

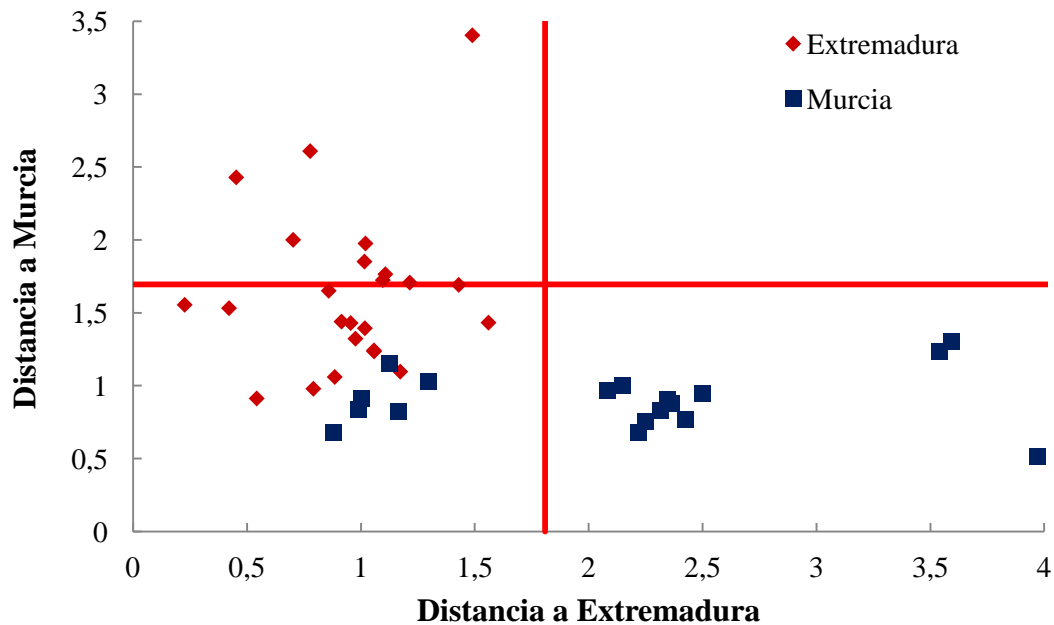


Figura IV.49.Gráfico de Cooman para las muestras del tipo picante según el modelo SIMCA^b.

Haciendo uso de redes neuronales artificiales del tipo perceptrón multicapa, MLP-ANN, se han conseguido mejores resultados. En el caso de MLP-ANN^a, la red se construye a partir de 6 neuronas en la capa de entrada, ya que son 6 variables las que se emplean para la construcción del modelo, 7 en la capa oculta y 2 en la capa de salida. En el caso de MLP-ANN^b la estructura de la red es 10:7:2, porque en este caso son las 10 variables, B, Ca, Cu, Fe, K, Mg, Mn, Pb, Sr y Zn las que se emplean para crear el modelo.

En ambos casos se tienen resultados mejores que los obtenidos utilizando otros métodos, por lo que la separación no lineal ofrece, efectivamente, mejores resultados. Considerando los resultados que se muestran en la tabla IV.19.se puede concluir que, para las muestras de pimentón picante la clasificación según el origen, no existen prácticamente diferencias a la hora de emplear las variables determinadas por Mann-Whitney o LDA^b para la construcción de las MLP-ANNs.

IV.2.2.3. Diferenciación geográfica de pimentones agridulces.

Se prepara una matriz de datos con 14 columnas, una por cada elemento analizado, y 45 filas correspondientes a cada una de las muestras de pimentón agridulce.

Se realiza un análisis PCA^a a la matriz de datos, utilizando como variables el contenido de los elementos Al, Ca, Fe, Na, Pb, Sr, Ni, Mn y Mg, seleccionados a partir de los resultados de la prueba U de Mann-Whitney. Con el contenido de estos descriptores se obtienen tres componentes principales (ec.4.40-4.42), que explican el 83,1% de la varianza de los datos originales. Los dos primeros explican el 70,2% de la misma, el PC1 un 50,1%, el PC2 un 20,1%, y el PC3 un 12,9%. Se consideran estos tres PCs, puesto que sus autovalores son mayores que 1: 4,51, 1,81 y 1,16 para PC1, PC2 y PC3, respectivamente.

$$\text{PC1: } 0,41|Al|+0,38|Fe|+0,41|Mn|+0,03|Sr|+0,32|Ni|-0,34|Pb| -0,10|Na| \\ +0,38|Mg|+0,38|Ca|(4.40)$$

$$\text{PC2: } -0,03|Al|+0,19|Fe|-0,05|Mn|+0,66|Sr|-0,17|Ni|+0,22|Pb| -0,61|Na| \\ -0,22|Mg|+0,16|Ca|(4.41)$$

$$\text{PC3: } -0,35|Al|-0,42|Fe|+0,33|Mn|+0,17|Sr|+0,43|Ni|+0,45|Pb| -0,21|Na| \\ -0,37|Mg|-0,05|Ca|(4.42)$$

La distribución de las muestras de pimentón agridulce según su origen MU y EX en el espacio definido por los tres primeros PCs se puede ver en la figura IV.50. Si consideramos las coordenadas de PC1, PC2 y PC3, las muestras de Murcia se distribuyen en la zona negativa de PC1 y PC2 y en la frontera de la zona positiva y negativa de PC3. Las muestras de Extremadura se distribuyen con valores positivos de PC1, negativos y positivos, en su mayoría, de PC2 y alrededor de todo el eje PC3. Atendiendo al peso de las variables originales en el primer componente principal, las que mayor poder discriminante tienen son Al, Ca, Fe, Mg, Mn y Pb, mientras que si observamos el PC2 tenemos que las variables que mayor contribución presentan son Sr y Na, y por último en el PC3, no existen variables con un peso superior a 0,7 pero algunas como Ni, Pb y Fe se aproximan y son las variables con mayor peso.

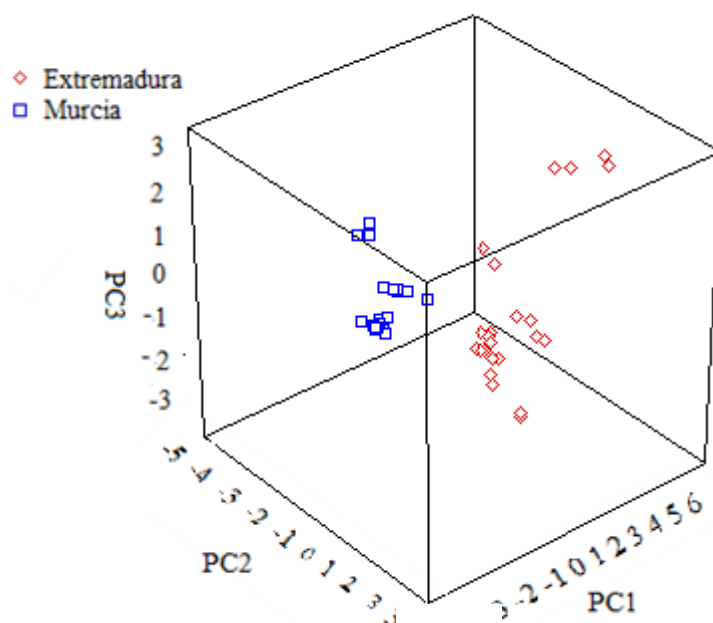


Figura IV.50. Distribución de muestras de pimentones agridulces según su origen en el plano definido por los tres primeros PCs.

En la tabla IV.20. se muestran los pesos de las variables en los distintos PCs

Tabla IV.20. Pesos de las variables en los PCs. (modelo PCA^a).
Pimentón agridulce

Elemento	PC1	PC2	PC3
Al	0,88	-0,04	-0,38
Ca	0,81	0,22	-0,01
Fe	0,81	0,25	-0,45
Mg	0,80	-0,29	-0,39
Mn	0,88	-0,07	0,35
Na	-0,20	0,82	-0,23
Ni	0,68	-0,23	0,47
Pb	0,71	0,29	0,48
Sr	0,07	0,89	0,19

El caso de PCA^b es diferente, ya que se parte de todas las variables. Con los tres primeros PCs, cuyas ecuaciones se incluyen en ec. 4.43-4.45, se explica el 73% de la varianza original. Se consideran los tres primeros PCs atendiendo al criterio de Kaiser. Los valores de los autovectores para PC1, PC2 y PC3 son 5,78, 2,93 y 1,50, respectivamente.

$$PC1: -0,30|Al| - 0,30|B| - 0,31|Ca| - 0,33|Cu| - 0,31|Fe| + 0,02|K| - 0,24|Mg| - 0,36|Mn| - 0,04|Na| - 0,25|Ni| + 0,09|P| - 0,34|Pb| - 0,17|Sr| - 0,34|Zn| \quad (4.43)$$

$$PC2: -0,29|Al| + 0,25|B| - 0,15|Ca| + 0,31|Cu| - 0,17|Fe| - 0,23|K| - 0,39|Mg| - 0,14|Mn| + 0,41|Na| - 0,20|Ni| - 0,11|P| + 0,04|Pb| + 0,43|Sr| + 0,27|Zn| \quad (4.44)$$

$$PC3: -0,10|Al| - 0,14|B| - 0,03|Ca| - 0,03|Cu| - 0,08|Fe| + 0,44|K| - 0,35|Mg| + 0,11|Mn| - 0,07|Na| + 0,20|Ni| - 0,71|P| + 0,28|Pb| - 0,03|Sr| - 0,09|Zn| \quad (4.45)$$

Como se puede observar en la tabla IV.21. los elementos que más contribuyen a PC1 son Al, B, Ca, Cu, Fe, Mn, Pb y Zn. Na y Sr presentan mayor contribución a PC2 y P a PC3.

Tabla IV.21. Peso de las variables en los PCs. (modelo PCA^b).
Pimentón agridulce

Elemento	PC1	PC2	PC3
Al	-0,73	-0,5	-0,12
B	-0,72	0,42	-0,18
Ca	-0,75	-0,25	-0,04
Cu	-0,79	0,53	-0,04
Fe	-0,74	-0,28	-0,10
K	0,04	-0,39	0,54
Mg	-0,57	-0,66	-0,43
Mn	-0,87	-0,25	0,14
Na	-0,09	0,71	-0,09
Ni	-0,6	-0,35	0,24
P	0,21	-0,19	-0,86
Pb	-0,81	0,07	0,35
Sr	-0,40	0,73	-0,04
Zn	-0,82	0,47	-0,11

La figura IV.51. muestra la distribución de las muestras de pimentón agridulce en el espacio definido por los tres primeros PCs. La mayoría de las muestras de MU presentan una distribución en la zona positiva de PC1 y PC2, mientras que las muestras de EX se distribuyen en torno a valores menores de PC2. Las muestras, tanto de MU como de EX se distribuyen a lo largo del eje PC3.

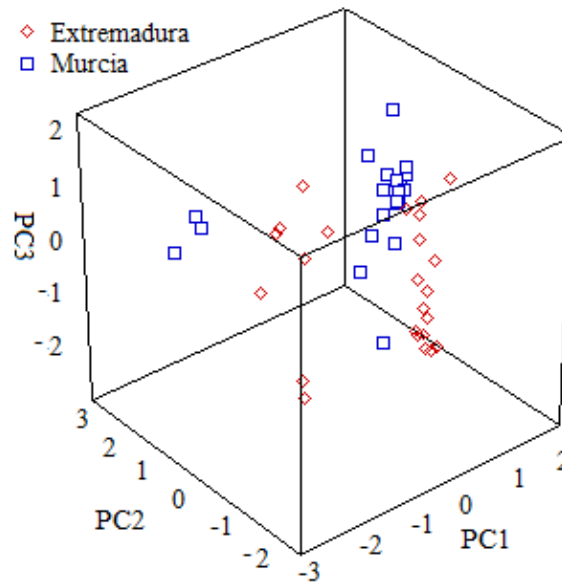


Figura IV.51. Distribución de muestras de pimentón agridulce según su origen en el plano definido por los tres primeros PCs.

Una vez realizado el PCA, se puede observar la tendencia natural de los datos a separarse. A continuación se aplica LDA para generar el modelo LDA^a. Se aplica un análisis sucesivo que incluye las variables Fe, Mn, Sr, Ni, Pb, Na, Mg y Ca. En la figura IV.52. se observa la separación que se consigue entre tipos de pimentones utilizando este modelo. En la ecuación 4.46 se presenta la función discriminante con la contribución de cada variable.

$$DF1: + 1,55|Fe| + 3,00|Mn| - 1,70|Sr| - 0,31|Ni| - 2,69|Pb| - 0,64|Na| + 0,23|Mg| + 0,69|Ca| \quad (4.46)$$

En la tabla IV.22. se muestran los resultados de la capacidad del modelo creado para diferenciar orígenes de muestras de pimentón agridulce en términos de SENS y SPEC. Aunque se obtenga un resultado del 100%, LDA es un método de modelado duro, con lo que puede que si se realiza una clasificación con SIMCA, la SENS y la SPEC disminuyan porque haya muestras que no se asignen a ninguna clase o que se asignen a ambas.

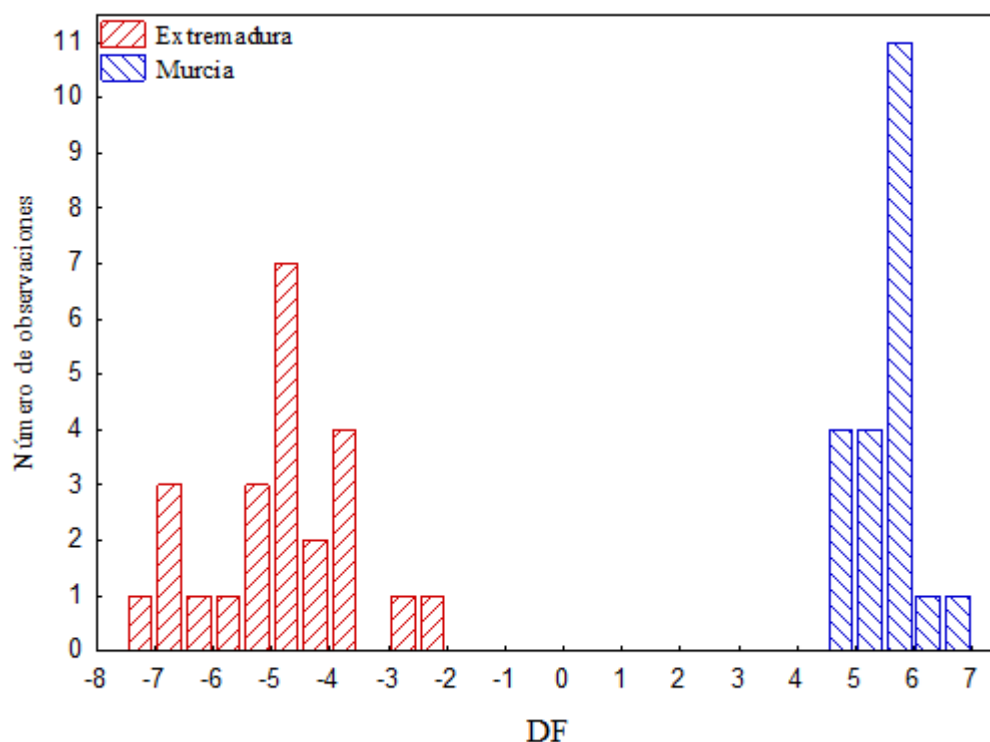


Figura IV.52. Distribución de las muestras de pimentones del tipo agridulce en DF según el origen, aplicando modelo LDA^a.

Tabla IV.22. Resultados de los modelos de clasificación global para muestras de pimentón agridulce.

	MU		EX		Global (MU-EX)	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA ^a	100	100	100	100	100	100
LDA ^b	100	100	100	100	100	100
SVM ^a	100	100	100	100	100	100
SVM ^b	100	100	100	100	100	100
SIMCA ^a	76 ± 17	100	96 ± 11	98 ± 7	93 ± 7	99 ± 3
SIMCA ^b	73 ± 24	100	91 ± 15	100	86 ± 12	100
MLP-ANN-10:8:2 ^a	100	100	100	100	100	100
MLP-ANN-12:8:2 ^b	100	100	100	100	100	100

MU: Murcia; EX: Extremadura; G: Global. a: Mann-Whitney. b: LDA con todas las variables y SVM, SIMCA y ANN con las variables incluidas al aplicar el modelo por LDA^b.

Siguiendo el procedimiento anterior, se genera un modelo alternativo de LDA empleando todas las variables de partida, LDA^b. Igualmente se aplica un análisis de inclusión sucesiva de variables y las únicas variables que no se incluyen en este caso

son el Al y el Mn. En la figura IV.53.se observa la separación que se consigue entre tipos de pimentones utilizando esas variables, y en la ecuación 4.47 se presenta la función discriminante con la contribución de cada variable.

$$DF1: -1,01|Mg| + 2,22|Sr| + 2,89|Pb| + 0,61|P| + 0,38|Ni| - \quad (4.47)$$

$$0,90|K| - 0,30|Ca| - 1,37|Fe| + 1,20|Na| - 2,00|Zn| - 0,82|B| + 1,43|Cu|$$

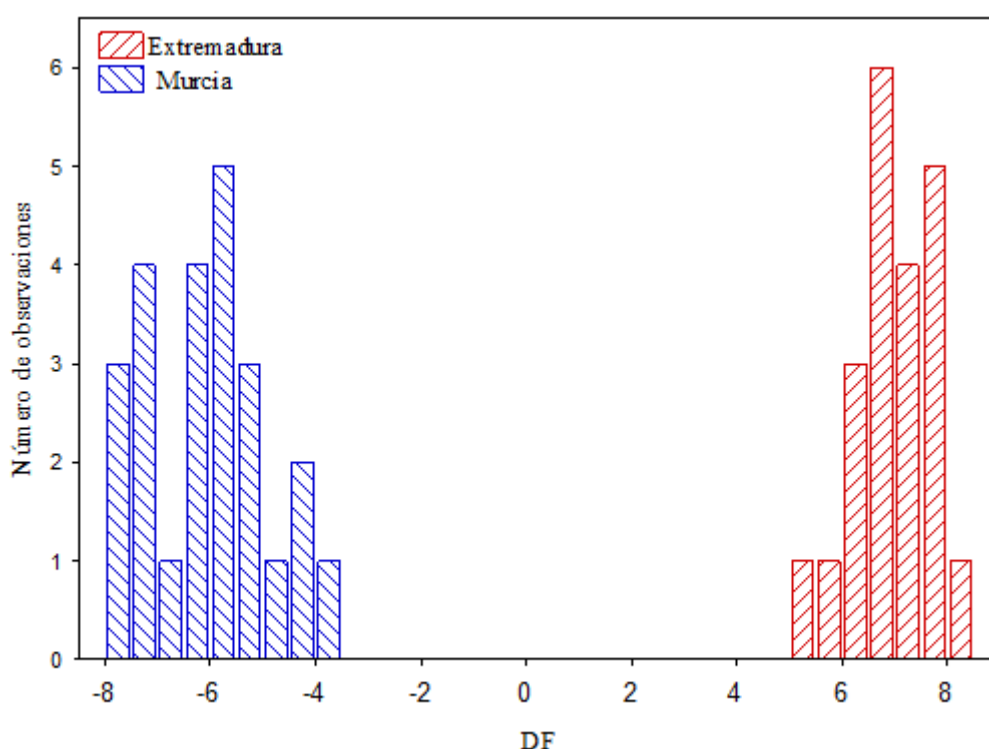
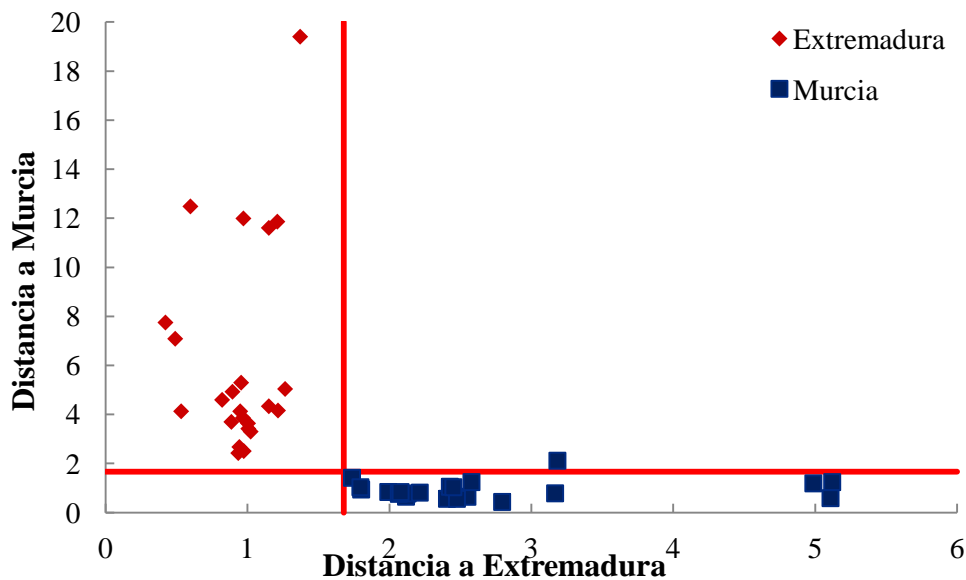


Figura IV.53. Distribución de las muestras de pimentones del tipo agridulce en DF según el origen, aplicando modelo LDA^b.

Se obtienen resultados del 100% en el caso del modelado mediante LDA^b, pero como se ha comentado anteriormente, habría que construir un modelo SIMCA para comprobar que no haya muestras que no se asignen a ninguna de las clases o a las dos.

Si se sigue con el esquema propuesto, el siguiente paso es construir dos modelos de clasificación mediante SVM. Uno de ellos usa todas las variables excepto B, Cu, K y Zn, que son las variables que Mann-Whitney considera con menor poder discriminante, SVM^a. Por otro lado, se crea el modelo SVM^b con las variables que LDA^b determinó como más determinantes, es decir, todas menos Al y Mn. En ambos casos, con los modelos SVM se consiguen resultados del 100%.

El modelo SIMCA^a, no proporciona resultados de SENS del 100%, ya que algunas muestras de MU no se asignan a ninguna clase, o se asignan erróneamente. En el gráfico de Cooman de la figura IV.54.se observa este hecho.



FiguraIV.54. Gráfico de Cooman para las muestras de pimentón agridulce según el modelo SIMCA^a.

En el caso del modelo SIMCA^b ocurre algo similar con los valores de SENS que disminuyen y se alejan del 100%. En este caso, la especificidad sí se mantiene en 100% porque no hay ninguna muestra erróneamente asignada a los dos orígenes. Los resultados de la tabla IV.22.se refieren a los resultados globales teniendo en cuenta los 9 replicados, y por tanto, aunque en el gráfico de Cooman (figura IV.55.) no se observen muestras de EX asignadas como no clase, la SENS no es del 100%.

En el caso de las muestras agridulces, también se recurre a técnicas no lineales como las MLP-ANN con fines comparativos.

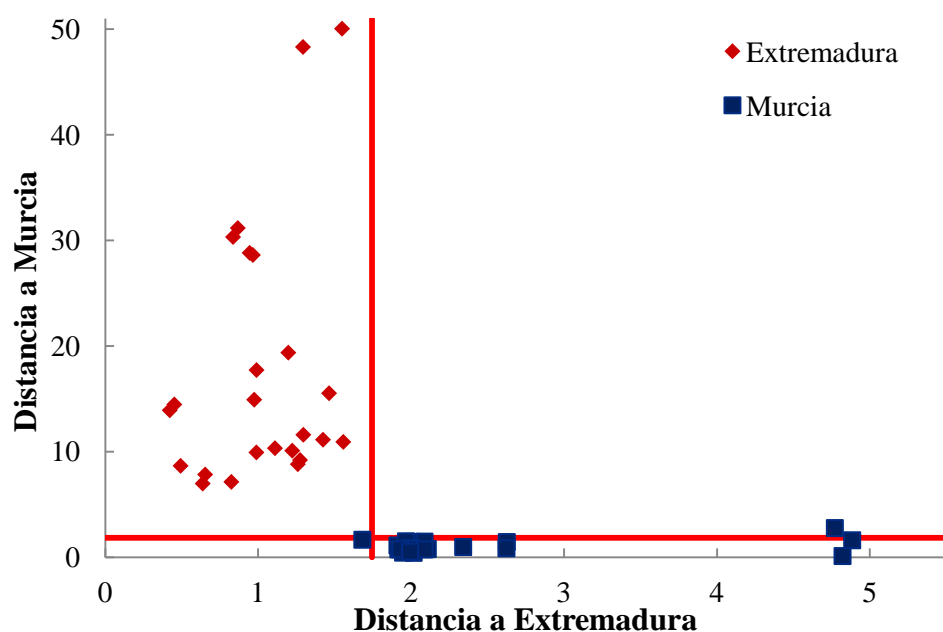


Figura IV.55. Gráfico de Cooman para las muestras de pimentón agridulce según el modelo SIMCA^b.

Se construyen dos modelos, MLP-ANN^a y MLP-ANN^b. En el primer caso las neuronas correspondientes a la capa de entrada son 10, mientras que en el segundo caso son 12. La capa oculta se compone de 7 y 8 neuronas, respectivamente, quedando dos modelos de arquitectura 10:8:2 y 12:8:2, para MLP-ANN^a y MLP-ANN^b, respectivamente. Los resultados se muestran en la tabla IV.22., y como se puede observar son del 100% en todos los casos, demostrando que el uso de una técnica no lineal permite la correcta clasificación de muestras de pimentón agridulce según origen de procedencia MU o EX.

IV.2.2.4. Diferenciación geográfica de pimentones mediante un modelo global

Una vez realizado el estudio sobre la diferenciación de muestras, de los tipos dulce, picante y agridulce, según su origen geográfico, se he planteado la opción de utilizar modelos que permitan diferenciar el origen geográfico de las muestras independientemente del tipo al que pertenezcan. Para comenzar el estudio sobre este modelo global, se considera, en primer lugar la prueba U de Mann-Whitney para determinar cuáles son las variables más relevantes. Los resultados obtenidos se

presentaron en la tabla IV.13., donde se muestra que las variables consideradas como más discriminantes cuando se consideran todos los tipos de pimentones en conjunto son Cu, Mg, Mn, Na, P, Pb y Sr.

Se ha seguido el mismo esquema de trabajo planteado en la figura IV.37. Por tanto, primer lugar se ha calculado PCA, realizando un análisis PCA^a empleando las variables anteriormente seleccionadas como más discriminantes, y después un análisis PCA^b empleando todas las variables.

En el caso de PCA^a, se han calculado los tres primeros PCs, cuyos autovalores son 2,73, 1,47 y 1,25 para PC1, PC2 y PC3, respectivamente. Estos tres primeros PCs explican un 77,85% de la varianza total y la contribución de cada variable se presenta en las ecuaciones 4.48-4.50.

$$PC1:0,29|Cu|+0,08|Mg|+0,21|Mn|-0,05|Na|+0,08|P|+0,17|Pb|+0,17|Sr|(4.48)$$

$$PC2:0,01|Cu|+0,06|Mg|+0,18|Mn|+0,52|Na|+0,01|P|+0,02|Pb|+0,19|Sr|(4.49)$$

$$PC3:0,01|Cu|+0,35|Mg|+0,001|Mn|+0,005|Na|+0,41|P|+0,14|Pb|+0,08|Sr| (4.50)$$

Cuando se representan las muestras en el plano de PC1 y PC2 (figura IV.56.) se observa que las muestras del origen EX se distribuyen en torno a valores negativos de PC2 y negativos y positivos de PC1, formando dos grupos de muestras. Las muestras pertenecientes al origen MU se sitúan en el cuadrante opuesto, es decir, en la zona de valores positivos de PC2 y positivos y negativos de PC1. Existen muestras de MU situadas en la zona de valores negativos de PC1 y PC2. La separación no es completa, pero se intuye una tendencia a la separación de las muestras en función del origen geográfico.

Las variables más contribuyentes a PC1 son Cu y Mn, mientras que para PC2 las que poseen mayor poder discriminante son Na y Sr, aunque con un peso menor de 0,7 como muestra la tabla IV.23. En el caso de PC3, P es la única variable con un peso mayor de 0,7. Sin embargo, el Mg también es un elemento que presenta una también una fuerte contribución.

Tabla IV.23. Peso de las variables en los PCs. (modelo PCA^a).

Glogal

Elemento	PC1	PC2	PC3
Cu	-0,90	0,12	-0,08
Mg	-0,45	-0,31	-0,66
Mn	-0,76	-0,51	-0,02
Na	-0,12	0,88	-0,08
P	0,45	-0,13	-0,72
Pb	-0,69	-0,15	0,42
Sr	-0,68	0,53	-0,32

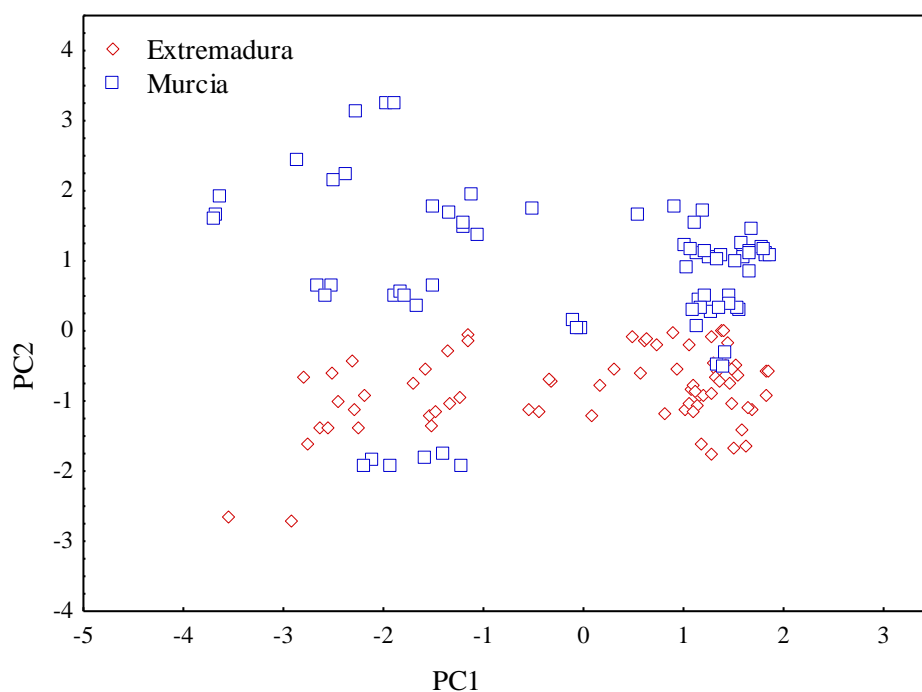


Figura IV.56. Distribución de las muestras según su origen en el plano de los dos primeros PCs.

Por otro lado, se ha realizado el cálculo del modelo PCA^b, utilizando todas las variables. Se obtienen 3 PCs, que explican el 64,7% de la varianza total. A continuación se considerará cuál es el peso de las variables en cada uno de los PCs calculados. Atendiendo a los valores incluidos en la tabla IV.24 se puede comprobar cuáles son los elementos que están más correlacionados con PC1, PC2 y PC3.

Tabla IV.24. Peso de las variables en los PCs. (modelo PCA^b).
Global

Elemento	PC1	PC2	PC3
Al	-0,78	-0,46	0,09
B	-0,69	0,33	-0,07
Ca	-0,76	-0,19	0,16
Cu	-0,87	0,25	-0,25
Fe	-0,70	-0,61	-0,14
K	0,23	-0,50	0,17
Mg	-0,58	-0,62	0,27
Mn	-0,79	0,27	0,40
Na	-0,05	-0,20	-0,77
Ni	-0,49	0,05	0,42
P	0,30	-0,11	0,16
Pb	-0,54	0,28	0,16
Sr	-0,64	-0,09	-0,64
Zn	-0,77	0,49	-0,09

Como se puede observar, PC1 está correlacionado fuertemente con variables como Al, Ca, Cu, Fe, Mn y Zn y moderadamente correlacionado con B, Mg, Pb y Sr. PC3 está correlacionado con el contenido de Na y moderadamente con Sr. Sin embargo, no se encuentran correlaciones importantes con PC2. La figura IV.57 muestra la distribución de las muestras en el espacio formado por los tres PCs. Se aprecian dos grupos de muestras correspondientes al origen EX, uno a valores negativos de PC1 y positivos de PC2, y otro grupo en la zona correspondiente a valores positivos de PC1 y de PC2. Las muestras correspondientes al origen MU presentan una dispersión mayor a lo largo del eje del PC2. Las muestras de este origen se distribuyen formando dos agrupaciones, que se sitúan a valores positivos y negativos de PC1. La mayoría de las muestras del origen EX se sitúan en la zona de valores positivos de PC3, mientras que las del origen MU presentan scores negativos en PC3, situándose en la parte inferior del gráfico, que corresponde a la parte negativa del eje PC3.

Utilizando LDA se han calculado dos tipos de modelos de clasificación. Uno de ellos utiliza las variables que con la prueba Mann-Whitney se consideraron como más discriminantes, LDA^a, y por otro lado usando todas las variables, LDA^b.

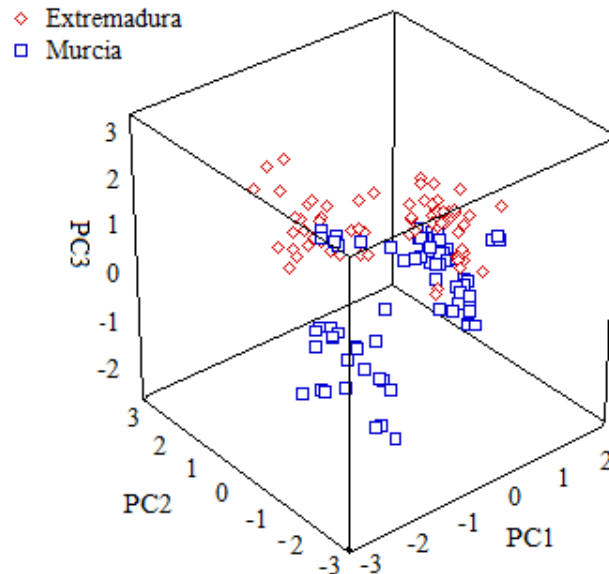


Figura IV.57. Distribución de las muestras según su origen en el plano de los tres primeros PCs.

Para el modelo LDA^a se aplicó un análisis LDA de inclusión progresiva para obtener las variables más discriminantes, que son Cu, Mn, Na, P, Pb y Sr. En la figura IV.58. se representan las muestras del origen EX distribuidas en la parte positiva de la DF, mientras que las muestras del origen MU lo hacen a valores negativos. Existe cierto solapamiento, lo que concuerda con el hecho de que los resultados de la clasificación en términos de SENS y SPEC no sean del 100%. En la tabla IV.25. se muestran los resultados correspondientes a LDA^a. Los resultados de SENS a la hora de clasificar muestras del origen MU son mejores que en el caso del origen EX, y lo contrario ocurre con la SPEC. Considerando los resultados de manera conjunta, se obtiene un 90% tanto para SENS como SPEC. En la ecuación 4.51 se presenta la función discriminante, DF, con la contribución de cada variable.

$$DF1: -0,97|Cu| + 1,04|Mn| - 0,37|Na| + 0,50|P| + 0,60|Pb| - 0,28|Sr| \quad (4.51)$$

El modelo LDA^b se obtiene a partir de un análisis de inclusión progresiva de las variables más discriminantes. En este caso el modelo se construye con todas las variables, excepto el K.

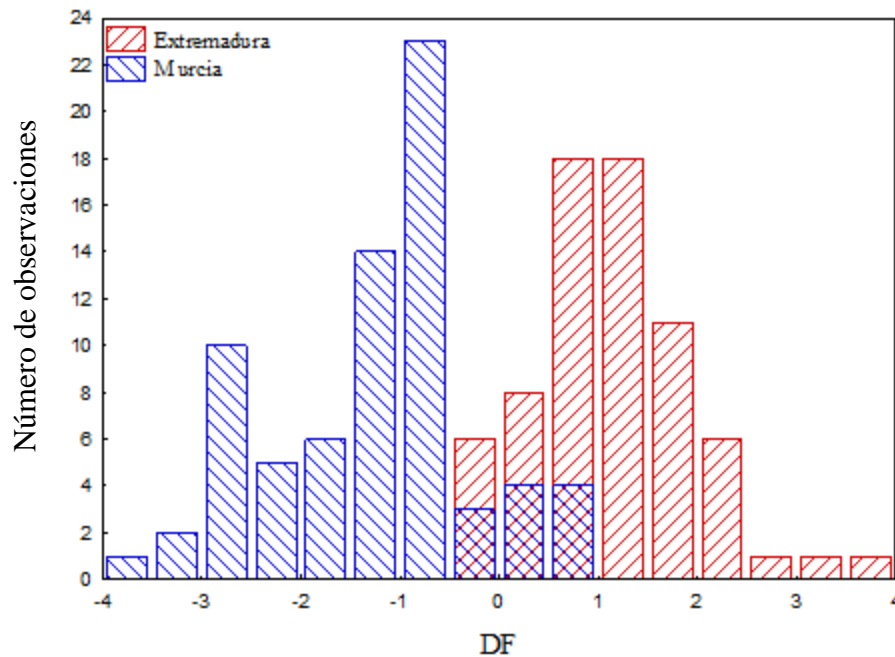


Figura IV.58. Distribución de las muestras de pimentón global en DF según el origen, aplicando modelo LDA^a.

En la figura IV.59. se ha incluido la distribución de las muestras de acuerdo con la función discriminante calculada según LDA^b. Se aprecia que las muestras correspondientes al origen EX se distribuyen en torno a valores negativos de DF1, mientras que las muestras correspondientes al origen MU lo hacen a valores positivos. También se observa cierto solapamiento en la zona central del gráfico. En la tabla IV.25. se muestran los resultados obtenidos con este modelo en términos de SENS y SPEC. Para el origen EX, los valores de SENS y SPEC son 98 y 94%, respectivamente, mientras que para el origen MU son 94 y 98%, respectivamente. Haciendo un cálculo promedio o global entre los dos orígenes se obtiene un valor del 96%, tanto para SENS como para SPEC. Este comportamiento se puede deber a las muestras del tipo picante, que, como ya se comprobó en el apartado IV.2.2.2 no se diferenciaban con los modelos LDA. No obstante, los valores de SENS y SPEC obtenidos con el modelo LDA^b permiten una diferenciación bastante aceptable.

Tabla IV.25. Resultados de los modelos de clasificación global para muestras de pimentones dulces, picantes y agridulces.

	MU		EX		Global (MU-EX)	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA ^a	87 ± 7	94 ± 4	94 ± 4	87 ± 7	90 ± 4	90 ± 4
LDA ^b	94 ± 7	98 ± 4	98 ± 4	94 ± 7	96 ± 4	96 ± 4
SVM ^a	98 ± 7	87 ± 4	87 ± 4	97 ± 7	93 ± 4	90 ± 4
SVM ^b	94 ± 7	87 ± 4	87 ± 4	94 ± 7	91 ± 4	90 ± 4
SIMCA ^a	27 ± 7	28 ± 15	60 ± 7	22 ± 17	43 ± 8	23 ± 6
SIMCA ^b	87 ± 6	64 ± 7	84 ± 7	84 ± 6	86 ± 3	74 ± 5
MLP-ANN-8:6:2 ^a	100	99 ± 3	99 ± 3	100	99 ± 1	99 ± 1
MLP-ANN-13:11:2 ^b	99 ± 4	99 ± 2	99 ± 2	99 ± 4	99 ± 2	99 ± 2

MU: Murcia; EX: Extremadura; G: Global. a: Mann-Whitney. b: LDA con todas las variables y SVM, SIMCA y ANN con las variables incluidas al aplicar el modelo por LDA^b.

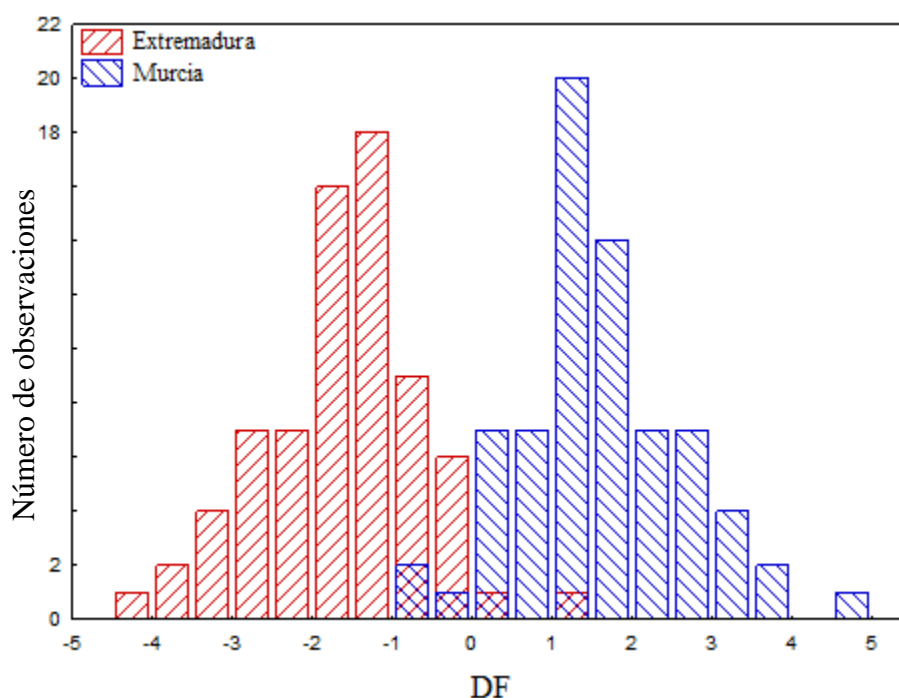


Figura IV.59. Distribución de las muestras de pimentones global en DF según el origen, aplicando modelo LDA^b.

Se han obtenido otros modelos de clasificación utilizando máquinas de vectores soporte, SVM. Al igual que con las otras técnicas se han realizado dos tipos de cálculo. SVM^a se ha obtenido utilizando las variables Cu, Mg, Mn, Na, P, Pb y Sr. Para el modelo SVM^b se han utilizado todas las variables excepto el K. La tabla IV.25. muestra los resultados obtenidos. Los valores de SENS y SPEC son del orden del 90%.

También se ha aplicado SIMCA, que tiene la ventaja de que permite visualizar las muestras asignadas a cada origen, y también cuáles no se clasifican de manera correcta. El modelo SIMCA^a se construye utilizando las variables que Mann-Whitney consideró con mayor poder discriminante, y SIMCA^b con las variables que quedaron en el modelo tras aplicar LDA^b.

En las figuras IV.60.y IV.61. se muestran los gráficos de Cooman de SIMCA^a y SIMCA^b, respectivamente. En ambos casos se puede observar cómo hay muchas muestras que se asignan a los dos orígenes geográficos. Cuando se calculan SENS y SPEC, para estos modelos, se obtienen valores especialmente bajos, que para SIMCA^a son inferiores al 50% (tabla IV.25.). En el caso de SIMCA^b los valores varían entre 64 y 87%.En los gráficos de Cooman se aprecia que hay bastantes muestras que no se asignan a ningún origen, y que también hay muestras asignadas a los dos orígenes.

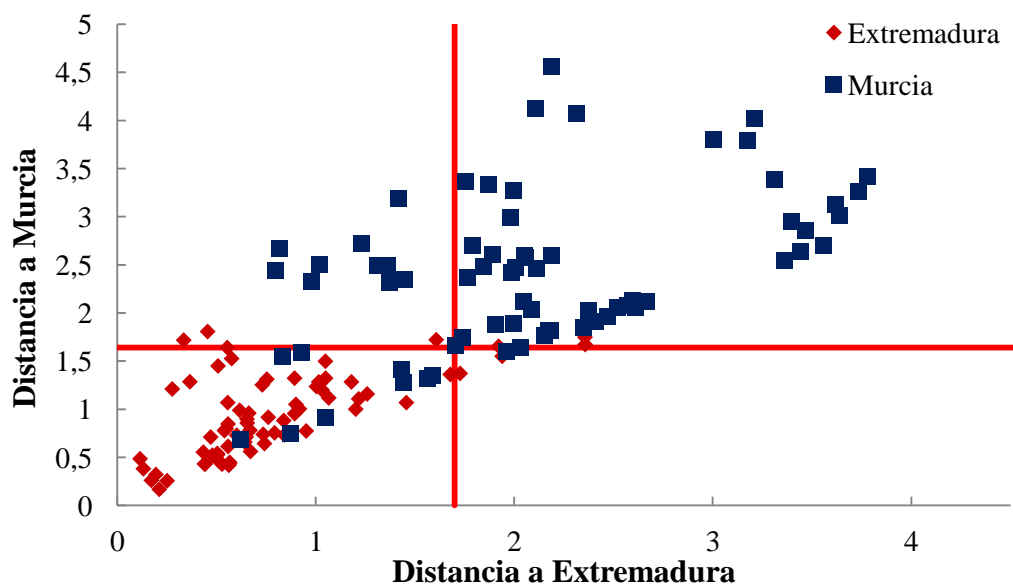


Figura IV.60. Gráfico de Cooman aplicando el modelo SIMCA^a. Global

Al aplicar tanto LDA como SVM se han obtenido mejores resultados que con SIMCA, lo cual puede significar que al usar una técnica lineal de modelado duro se enmascaren resultados, y por ello se consiguen porcentajes más altos. El uso de técnicas no lineales puede permitir una asignación más fiable de las muestras de pimentón según

su origen geográfico, sin tener en cuenta los tipos dentro de cada origen. Por ello, y con fines comparativos se han utilizado modelos de clasificación basados en redes neuronales artificiales.

En las clasificaciones anteriores fue necesario el uso de MLP-ANN para conseguir resultados adecuados de SENS y SPEC. En el caso de los pimentones dulces y agridulces, con las MLP-ANN se consiguieron resultados del 100% de SENS y SPEC. Sin embargo, para la clasificación de pimentones picantes el uso de esta técnica no lineal no proporcionó resultados del 100%, aunque sí mejoraron los obtenidos con modelos lineales. Para la clasificación global cabe esperar, por tanto, que los resultados no sean del 100% ya que las muestras picantes introducen ese error a la hora de la clasificación.

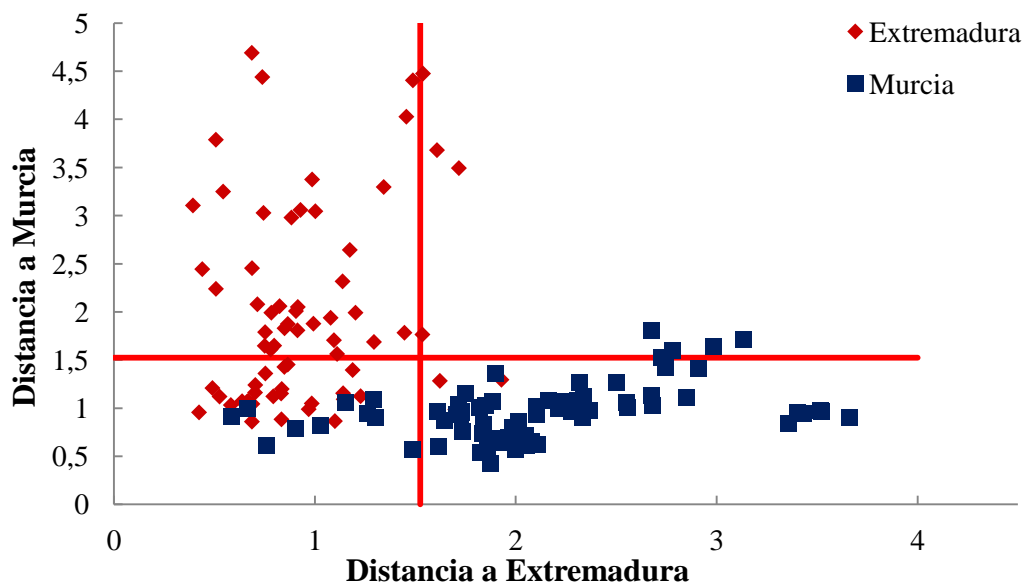


Figura IV.61. Gráfico de Cooman aplicando el modelo SIMCA^b. Global

Se generan dos modelos MLP-ANN al igual que en los casos anteriores; en primer lugar MLP-ANN^a con las variables, Cu, Mg, Mn, Na, P, Pb y Sr; en segundo lugar MLP-ANN^b con Al, B, Ca, Cu, Fe, Mg, Mn, Na, Ni, P, Pb, Sr y Zn.

El primer modelo de MLP-ANN^a presenta una arquitectura 8:6:2, y los resultados, tanto de SENS como de SPEC, mejoran respecto a modelos lineales,

obteniéndose valores globales de SENS y SPEC mayores al 99% (tabla IV.25.). El modelo MLP-ANN^b utiliza una red con 13 neuronas en la capa de entrada, 11 en la capa oculta y 2 en la de salida. SENS y la SPEC, tal y como se muestra en la tabla IV.25., presentan valores del 99%, mejorando considerablemente los resultados obtenidos tras aplicar métodos lineales.

IV.3. Diferenciación geográfica de pimentón mediante el estudio del color y espectros de absorción ultravioleta visible.

En este apartado se propone una metodología rápida y económica para la diferenciación de origen de muestras de pimentón de las denominaciones de origen de Murcia y de la Vera. El problema de la diferenciación se va a abordar desde dos puntos de vista, por un lado se calculan los parámetros relacionados con el color, ASTA y las coordenadas CIELab, y por otro se hace un estudio considerando el espectro de absorción en el rango UV-Vis. Se trata de un procedimiento sencillo que no requiere instrumentación sofisticada, con una preparación muy simple de las muestras [27].

Se registra el espectro de absorción en el rango de longitud de onda de 250-800 nm de un extracto de pimentón en acetona. A partir de los datos de absorbancia se calcula el parámetro ASTA y las coordenadas L* a* y b*. Con el fin de realizar la diferenciación geográfica se aplican técnicas como PCA, LDA, SVM o ANN. En primer lugar, para poder observar las tendencias que presentan los datos a diferenciarse, se aplica PCA. A continuación, se construyen modelos lineales como LDA o SVM, y modelos no lineales como SVM-no lineales y ANN, para los casos en que la diferenciación no sea posible con un modelo lineal. Por otro lado, se lleva a cabo la diferenciación mediante una segunda estrategia utilizando los datos de absorbancia de los espectros transformados en sus correspondientes PC-scores mediante PCA. Esta transformación permite la reducción del número de variables a utilizar en los modelos de clasificación, ya que no es viable trabajar con todo el rango de longitudes de onda y además son variables muy correlacionadas entre sí. A los PC-scores, de manera análoga a como se hizo con las coordenadas del color, se aplica, en primer lugar, un modelo LDA. Si LDA no genera resultados adecuados, se emplean SVM o ANN.

En todos los casos, se divide la matriz de datos en un subconjunto de entrenamiento (75% de las muestras) que se emplea para construir el modelo de clasificación, y un subconjunto de prueba (25% de las muestras) para calcular la eficacia del mismo. La eficacia se evalúa a través del cálculo de los parámetros sensibilidad (SENS) y especificidad (SPEC) empleando las ecuaciones 4.25 y 4.26. Para realizar la validación cruzada del modelo se lleva a cabo un procedimiento de submuestreo DAGJK con nueve replicados, similar al realizado en el epígrafe dedicado al estudio del perfil metálico. Cuando se calculan modelos de clasificación mediante redes neuronales de perceptrón multicapa, MLP-ANN, es necesario incluir un tercer set de muestras llamado de verificación para evitar el sobreentrenamiento, ya que estos modelos se entrenan mediante retropropagación (back propagation, BP). Se emplea el mismo método aleatorio DAGJK, pero el set de entrenamiento se divide en un 50% para verificación y un 50% para entrenamiento.

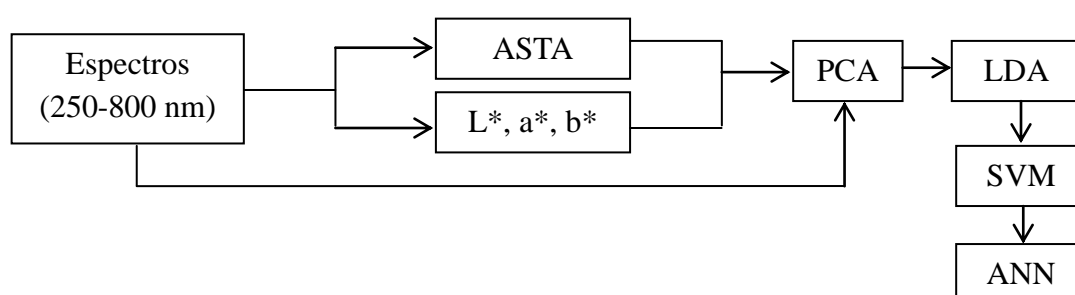


Figura IV.62. Esquema del proceso de obtención de modelos de clasificación.

IV.3.1. Estudio del color en base al parámetro ASTA y coordenadas CIELab para la diferenciación de pimentones.

En este epígrafe se presenta el estudio del color de extractos de pimentones de acuerdo con el método ASTA [31] y el procedimiento de obtención de coordenadas CIELab [32].

El parámetro ASTA puede obtenerse de la medida de la absorbancia a 460nm de acuerdo al procedimiento descrito en el apartado III.2.3.1.

En la tabla A1.2 del anexo 1 se incluyen los resultados de los valores ASTA para todas las muestras de los distintos tipos. Para poder analizar de manera más sencilla los datos, se incluyen los diagramas de caja y bigotes (figura IV.63.) para la variable ASTA, así como una tabla en la que están recogidos los valores de las medianas y los máximos y mínimos de ASTA para los distintos orígenes y tipos (tabla IV.26.).

En el caso de los pimentones tipo dulce, los valores más altos de ASTA se corresponden al origen Extremadura, siendo 140 y 124 los valores máximos para Extremadura y Murcia, respectivamente. El rango de los valores ASTA para el tipo dulce es de 83-140 y 63-124, para Extremadura y Murcia, respectivamente. El valor de la mediana es mayor en el caso de las muestras de Extremadura.

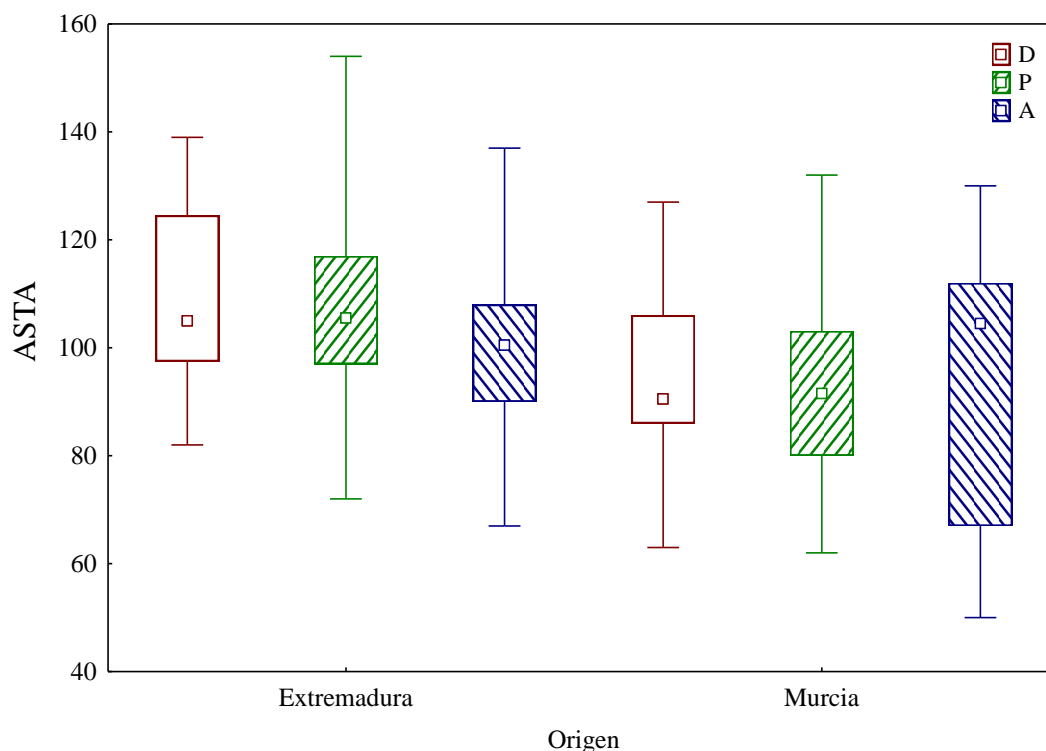


Figura IV.63. Diagrama de caja y bigotes del parámetro ASTA para muestras de pimentón dulce (D), picante (P) y agridulce (A).

Las muestras de pimentón picante del origen Extremadura son los que presentan mayores valores de ASTA, presentando dicha procedencia un valor máximo de 155. Las diferencias entre los pimentones picantes de Extremadura y Murcia se pueden observar en el diagrama de caja y bigotes de la figura IV.63. Los valores la mediana son parecidos si se comparan ambos orígenes, aunque aparecen diferencias entre los rangos obtenidos para ambas clases. En el caso de Extremadura, los valores de ASTA oscilan entre 73 y 155, mientras que para los pimentones procedentes de Murcia los valores se mueven en la horquilla de 62-119. Se observa además una mayor dispersión para los valores de este parámetro en el caso de las muestras de pimentón procedentes de Extremadura.

En el caso de los pimentones agridulces, cuando se realiza el estudio del parámetro ASTA, se observa una tendencia opuesta a la que se ha visto para los pimentones dulces y picantes. El valor de la mediana es mayor para las muestras procedentes de Murcia, aunque, los valores máximos se corresponden con el origen Extremadura. En el caso de los valores mínimos existe una diferencia más notable entre orígenes, siendo los pimentones de Murcia los que presentan valores menores. Tanto en la tabla IV.26. como en la figura IV.63., se puede observar estos hechos así como una mayor dispersión para las muestras procedentes de Murcia. En cuanto a los valores de la mediana, existe una diferencia más pronunciada entre los valores ASTA de los distintos orígenes en el tipo de pimentón dulce que en el agridulce. En el diagrama de caja y bigotes de la figura IV.63. se observan claramente estas diferencias en cuanto a mediana y valores máximos y mínimos.

Tabla IV.26. Valores de mediana, mínimo y máximo para el parámetro ASTA y las coordenadas L*, a*, b*, C* y H para los distintos orígenes y tipos de pimentón.

Variable	MURCIA			EXTREMADURA		
	Dulce	Picante	Agridulce	Dulce	Picante	Agridulce
ASTA	90,5	99,5	104,5	104,5	105,0	100,5
	(63-124)	(62-119)	(54-130)	(83-140)	(73-155)	(67-137)
L*	68,9	66,4	80,4	61,2	54,8	76,1
	(37,6-88,1)	(44,7-81,4)	(21,9-88,8)	(36,7-82,8)	(21,0-85,8)	(60,0-87,7)
a*	13,5	12,0	12,2	17,0	15,0	17,6
	(7,5-17,3)	(3,1-18,5)	(5,2-20,2)	(4,5-22,2)	(1,0-34,7)	(9,0-21,6)
b*	65,2	62,1	75,3	77,3	73,6	83,2
	(59,6-81,0)	(24,8-84,3)	(42,4-84,7)	(31,4-97,00)	(10,2-109,8)	(64,5-104,0)
C*	66,6	63,0	76,7	79,1	75,1	85,6
	(60,9-82,8)	(25,0-86,3)	(43,8-87)	(31,7-99,5)	(10,3-115,1)	(65,1-106,2)
H	66,6	63,0	79,9	77,7	77,3	78,2
	(76,4-83,3)	(76,8-83,4)	(75,4-84,2)	(76,5-82,0)	(72,4-84,4)	(76,1-82,0)

Valores mínimos y máximos entre paréntesis.

A partir del espectro de absorción en el intervalo 380-780 nm se calculan los valores triestímulo, que a su vez sirven para medir el color percibido y relacionarlo con los términos adecuados que describen el espacio de color CIELab. En primer lugar, a partir de las absorbancias (A) se obtienen los valores de transmitancia como:

$$T = 10^{-A} \quad (4.52)$$

Posteriormente se calculan los valores triestímulo, X, Y, Z según las ecuaciones 4.55-4.57.

$$X = K \cdot \sum_{\lambda} T_{(\lambda)} S_{(\lambda)} \bar{X}_{10(\lambda)} \Delta_{(\lambda)} \quad (4.53)$$

$$Y = K \cdot \sum_{\lambda} T_{(\lambda)} S_{(\lambda)} \bar{Y}_{10(\lambda)} \Delta_{(\lambda)} \quad (4.54)$$

$$Z = K \cdot \sum_{\lambda} T_{(\lambda)} S_{(\lambda)} \bar{Z}_{10(\lambda)} \Delta_{(\lambda)} \quad (4.55)$$

donde $T_{(\lambda)}$ es la medida de transmitancia de la muestra para cada longitud de onda, $\Delta_{(\lambda)}$ el paso de longitud de onda elegido para registrar el espectro de absorbancia en el rango de medida, en este caso 1nm, $S_{(\lambda)}$ es un coeficiente que depende del iluminante usado, $\bar{X}_{10(\lambda)}$, $\bar{Y}_{10(\lambda)}$ y $\bar{Z}_{10(\lambda)}$ son coeficientes que dependen del observador considerado y que están recogidos en las tablas publicadas por la CIE [33], y K es una constante de normalización que se calcula según [34]:

$$K = \frac{100}{\sum_{\lambda} S_{(\lambda)} \bar{Y}_{10(\lambda)} \Delta_{(\lambda)}} \quad (4.56)$$

Una vez obtenidos los valores triestímulo, es posible calcular a partir de ellos las coordenadas L^* , a^* y b^* , que definen el espacio CIELab, según las siguientes ecuaciones:

$$L^* = 116 \cdot F(2) - 16 \quad (4.57)$$

$$a^* = 500 [F(1) - F(2)] \quad (4.58)$$

$$b^* = 200 [F(2) - F(3)] \quad (4.59)$$

donde $F(1)$, $F(2)$ y $F(3)$ son funciones de los valores triestímulo, X, Y y Z, respectivamente. Para su cálculo se definen las funciones $G(i)$ siendo:

$$G(1) = X/X_n \quad (4.60)$$

$$G(2) = Y/Y_n \quad (4.61)$$

$$G(3) = Z/Z_n \quad (4.62)$$

donde X_n , Y_n y Z_n son las coordenadas triestímulo referencia de un blanco o cero, asumido como difusor perfecto para el intervalo visible del espectro 380-780 nm. Estos valores se obtienen de la misma manera que X, Y, Z, pero cuando la transmitancia es 1.

El uso del algoritmo en las ecuaciones (4.57-4.59) permite la unificación en una sola ecuación, teniendo en cuenta la variabilidad del factor luminancia, Y/Y_n [35],

$$\text{Si } G(i) > 0,008856 \Rightarrow f(i) = G(i)^{1/3} \quad (4.63)$$

$$\text{Si } G(i) < 0,008856 \Rightarrow f(i) = 7,787 \cdot G(i)^{1/3} + 16/116 \quad (4.64)$$

Para terminar, según proceda, se sustituye el valor de $F(i)$ en las ecuaciones 4.57-4.59.

El croma y el tono se pueden calcular según las siguientes expresiones [35]:

$$C^* = (a^{*2} + b^{*2})^{1/2} \quad (4.65)$$

$$H = \arctg\left(\frac{b^*}{a^*}\right) \quad (4.66)$$

En la tabla IV.26. se recogen los valores de mediana, mínimo y máximo para las coordenadas L^* , a^* y b^* , así como para C^* y H dependiendo del tipo y procedencia del pimentón. Los valores obtenidos para cada muestra se pueden consultar en la tabla A.1.2. del Anexo 1. Los términos C^* y H no se emplearán como variables para la diferenciación de orígenes ya que al obtenerse a partir de los valores a^* y b^* presentan una alta correlación con estos y no aportan información nueva al modelo. Además, la elevada correlación complica la obtención de modelos mediante LDA, pues se obtienen matrices singulares que no pueden invertirse.

Al igual que con los valores de ASTA, es posible comparar los resultados entre orígenes por tipos de pimentones para ver si existen algunas diferencias que indiquen que variables pueden emplearse como variables discriminantes en modelos posteriores. En este caso se representa en las figuras IV.64-IV.66. la distribución de las muestras en el espacio CIELab para estudiar las coordenadas L^* , a^* y b^* de los distintos tipos según el origen.

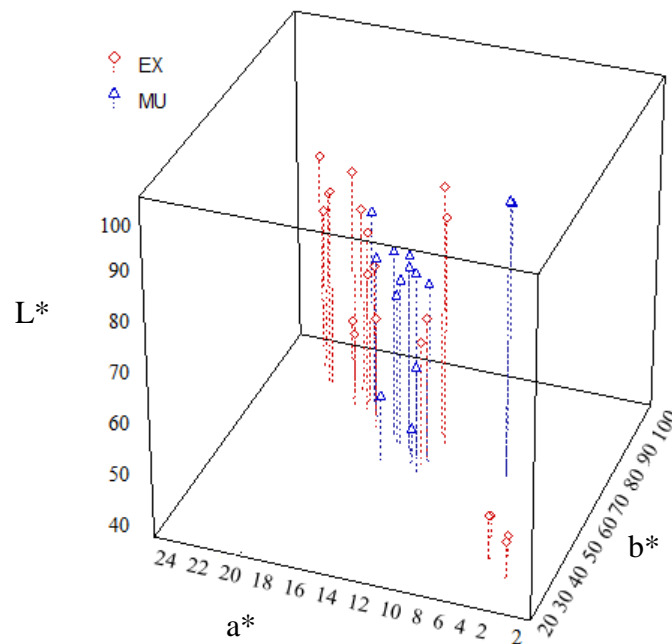


Figura IV.64. Distribución de las muestras de pimentones dulces en el espacio formado por L^* , a^* y b^* . EX: Extremadura, MU: Murcia.

Para las muestras del tipo dulce, se puede observar en la figura IV.64. que los valores de L^* son similares entre ambos orígenes, aunque algunas muestras procedentes de Murcia presentan valores más altos. En el caso de la coordenada a^* se observa como las muestras del origen Extremadura tienden a agruparse en torno a valores mayores que las muestras de Murcia. Aunque el valor máximo de b^* se corresponde con el origen Extremadura, existen unas muestras de dicho origen, situadas en la parte de valores inferiores de b^* .

En el caso de las muestras picantes, tal y como se observa en la figura IV.65. las muestras de Extremadura se distribuyen en torno a valores más altos de a^* y b^* . En cuanto a la coordenada L^* , aunque el valor máximo corresponda al origen Extremadura, la mayoría de las muestras de Murcia se sitúan en torno a valores más altos de L^* que las muestras de Extremadura. La dispersión de a^* dentro de las muestras de Extremadura es mayor que en el caso de Murcia, tal y como se puede observar en la figura.

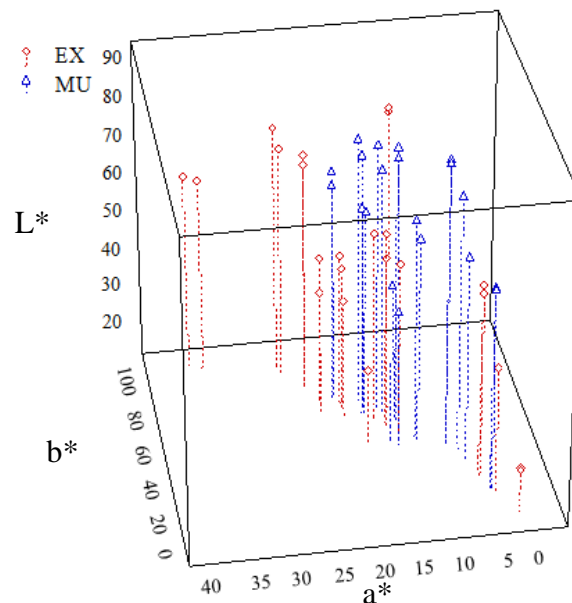


Figura IV.65. Distribución de las muestras de pimentones picantes en el espacio formado por L^* , a^* y b^* . EX: Extremadura, MU: Murcia.

Para las muestras pertenecientes al tipo agridulce (figura IV.67.), encontramos que las que proceden del origen Extremadura se agrupan, en su mayoría, en torno a valores superiores de a^* y de b^* . Las muestras de Murcia presentan mayor dispersión para el parámetro L^* .

En relación al conjunto total de muestras, con respecto al parámetro L^* se puede observar que los valores oscilan entre 21 y 88,85, con lo que la claridad varía bastante entre tipos y orígenes. En las figuras IV.65. se observa cómo los valores de L^* son menores para las muestras picantes y dulces procedentes de Extremadura, mientras que para el caso del tipo agridulce, encontramos los valores mínimos de L^* para las muestras de Murcia. En el caso de los pimentones agridulces hay diferencias importantes si se comparan los valores mínimos de los dos orígenes, aunque la diferencia no es tan llamativa si se comparan los valores medios. Para el caso de los pimentones dulces no encontramos tal diferencia, ya que el rango de valores es muy parecido. Sin embargo, si se comparan los resultados de L^* correspondientes a los pimentones picantes, se observa que el valor mínimo para los pimentones del origen Extremadura es mucho menor que el mínimo de los de Murcia siendo estos valores 21,04 y 44,68, respectivamente. No sólo los valores mínimos, sino también la media de

L^* para pimentones picantes es lo que los convierte en los de apariencia menos clara a la vista (figura IV.67.), ya que L^* es el parámetro que indica la dimensión de la claridad.

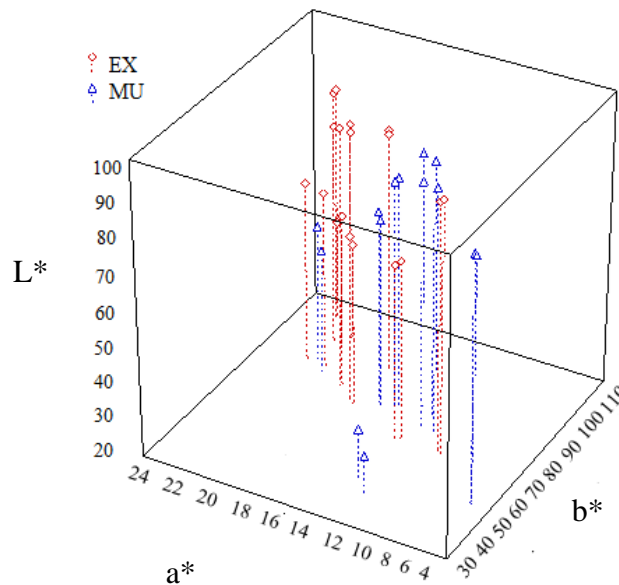


Figura IV.66. Distribución de las muestras de pimentones agridulces en el espacio formado por L^* , a^* y b^* . EX: Extremadura, MU: Murcia.

Las coordenadas a^* y b^* definen el plano de cromaticidad cuyo centro es neutral. La coordenada a^* da información sobre el rango rojo-verde ($a^* > 0$ rojo, $a^* < 0$ verde), y $b^* < 0$ es la dirección del azul, mientras que $b^* > 0$ es la dirección del amarillo [36]. En el plano formado por a^* y b^* llama la atención el caso de los pimentones picantes, que presentan los valores mínimos de a^* y de b^* , y en particular los valores de a^* y b^* del origen Extremadura, que son 1,0 y 10,2, respectivamente. Una disminución del valor del parámetro a^* indica que el color se torna menos rojo, y cuando el parámetro b^* es menor indica que se aleja del amarillo. En la figura IV.67.se aprecia como el color del pimentón picante de Extremadura se percibe como menos claro y menos rojo que el resto, así como los agridulces parecen ser los más claros y los más rojos. Los pimentones picantes de ambos orígenes muestran los valores más bajos de b^* siendo los valores de la mediana de b^* para los orígenes Extremadura y Murcia 73,6 y 62,1, respectivamente. Sin embargo, los valores máximos de b^* de los dos orígenes y los tres tipos son parecidos, siendo menores los del pimentón dulce.

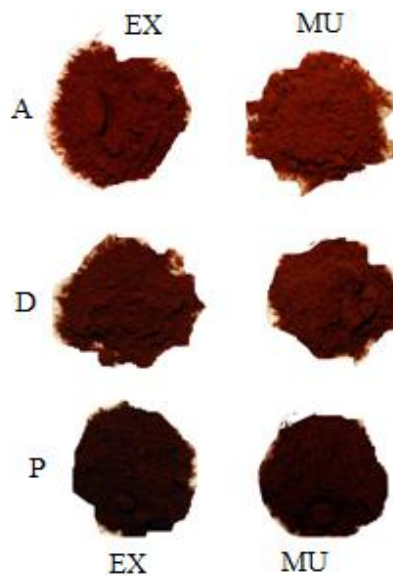


Figura IV.67. Muestras de pimentones de los tipos agridulces (A), dulce (D) y picantes (P) de los orígenes Extremadura (EX) y Murcia (MU).

El croma, C^* , que da cuenta de la intensidad del color puede tomar valores desde 0 a 150. En este caso los resultados se encuentran dentro del rango de la horquilla 23 a 115 (tabla IV.26.), siendo los pimentones agridulces los que mayores valores presentan, seguidos de los picantes. Para el cálculo del tono H, el signo de a^* y b^* es decisivo para obtener correctamente la posición de la coordenada H en el cuadrante adecuado. En este caso, como a^* y b^* son positivos, y por tanto trabajamos en el primer cuadrante (rojo-amarillo). El valor que obtenemos al aplicar la ecuación 4.66, oscila entre $0-90^\circ$, como es de esperar.

Considerando datos bibliográficos se encuentran concordancias con otros resultados publicados. En primer lugar, con respecto al ASTA, tanto el BOE que rige las características del pimentón de la Vera [37] como la norma referida al pimentón de Murcia [38], exige que el ASTA sea mayor de 90 unidades en el momento de la molienda. En algunos casos encontramos valores inferiores de ASTA pero esto es debido al periodo de almacenamiento. Topuz et al. [39] estudiaron el efecto que produce el secado y almacenamiento en el color de distintos tipos de pimientos, concluyendo que el parámetro ASTA puede disminuir hasta un 15% de los valores originales, pasando de un producto rojizo brillante a uno más marrón y apagado [40]. En *Spices and Seasonings: A Food Technology Handbook* [41] encontramos que los valores

típicos para los pimientos secados y molidos están en torno a 85-150 unidades ASTA. Los parámetros L^* , a^* y b^* nos proporcionan una información adicional a la que se obtiene tras estudiar la tendencia de los valores ASTA, ya que ofrecen información sobre la fracción rojo/amarillo. El parámetro ASTA da cuenta del contenido total de carotenoides y no es posible saber si un pimentón es más o menos rojo únicamente en función de su valor ASTA [42]. Los valores de L^* , a^* , b^* , C^* y H , están acordes a los valores publicados por otros autores, siendo a^* y b^* siempre valores positivos para el caso de muestras rojas, amarillas o anaranjadas [43]. Si se tratara con variedades de pimientos verdes las coordenadas, sobre todo la a^* , serían muy diferentes, ya que los valores negativos a^* son los relacionados con el color verde [44].

IV.3.1.1. Diferenciación geográfica de muestras de pimentones dulces a partir del parámetro ASTA y coordenadas CIELab.

En este apartado se incluyen los resultados del estudio de diferenciación de los orígenes de muestras de pimentón a partir de cuatro variables, como son el parámetro ASTA y las tres coordenadas que definen el espacio CIELab. En el esquema de trabajo a seguir se usará PCA como técnica de previsualización de la tendencia de los datos y, a continuación, se construirán modelos de clasificación a partir de LDA, SVM o ANN (figura IV.62.). En todos los casos se subdivide la matriz de datos y se aplica DAGJK como procedimiento de validación cruzada.

Cuando se realiza el PCA, se encuentra que con los dos primeros PCs se explica hasta un 96,39% de la varianza. En este caso los dos primeros PCs presentan autovalores mayor que la unidad, 2,32 y 1,54, respectivamente. Tanto PC1 como PC2 están muy correlacionados con las cuatro variables, en particular PC1 con L^* , a^* y b^* y PC2 con el parámetro ASTA, como se observa en la tabla IV.27.

Tabla IV.27. Peso de las variables en los PCs.

	PC1	PC2	PC3	PC4
ASTA	0,43	0,87	0,23	-0,01
L	-0,82	-0,53	0,23	-0,04
a	-0,77	0,61	-0,15	-0,07
b	-0,93	0,36	0,02	0,09

En la figura IV.68. se puede observar la distribución de las muestras en el plano formado por PC1 y PC2. Las muestras de MU y EX, casi en su totalidad, se distribuyen de forma dispersa en dicho plano, no observándose ninguna tendencia de separación clara.

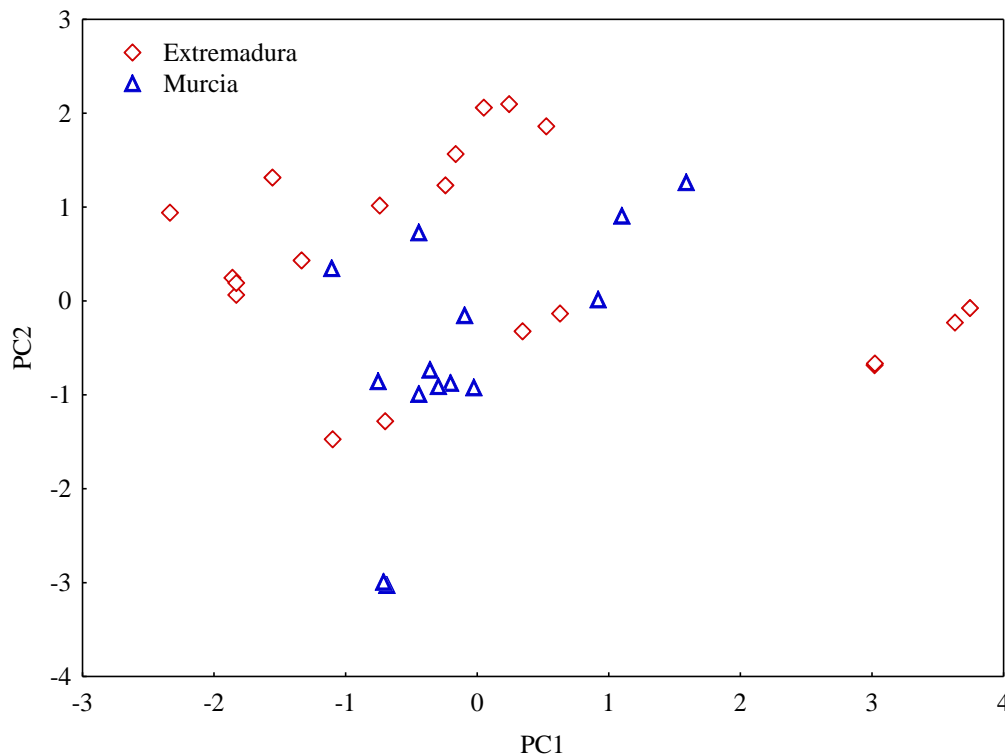


Figura IV.68. Distribución de las muestras de pimentón dulce en el espacio formado por PC1 y PC2.

Como primer paso para la creación de un modelo capaz de diferenciar orígenes, se aplica LDA. En este caso se obtendrán dos modelos, uno incluyendo todas las variables y un segundo seleccionando solo las más discriminantes.

El primer modelo se obtiene aplicando un LDA estándar utilizando las cuatro variables L^* , a^* , b^* y ASTA. En la figura IV.69. se puede apreciar cómo la separación no es completa, observándose solapamiento entre los dos orígenes en la zona donde la DF toma valores entre -1 y 1. En la tabla IV.28. se muestran los resultados de SENS y SPEC que son del orden del 70 y 65%, respectivamente.

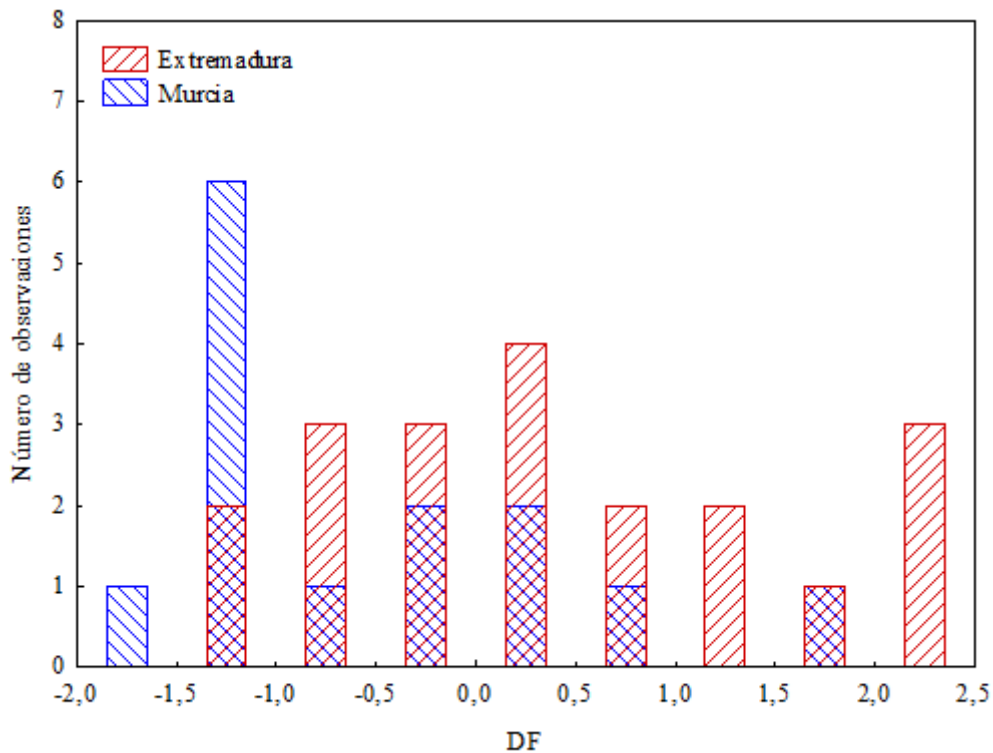


Figura IV.69. Distribución de las muestras de pimentones dulces en DF según el origen.

El segundo modelo que se construye emplea un LDA de inclusión progresiva (forward LDA, FLDA), donde las variables se van incluyendo en función de su poder discriminante. En este caso, cuando se emplea FLDA, encontramos que sólo considera que la variable ASTA es apropiada para la diferenciación, pero no es posible construir un modelo de clasificación con sólo una variable.

El siguiente paso es obtener modelos de clasificación mediante SVM. Se crean dos modelos con las cuatro variables, uno con Kernel del tipo función de base radial, RBF-SVM y otro con Kernel tipo sigmooidal, S-SVM. Los resultados globales no mejoran con respecto a LDA, y se encuentra, sobre todo en S-SVM, el problema de que la desviación estándar es demasiado elevada. En la tabla IV.28. se muestran estos resultados. La SENS para las muestras de MU en el caso de las SVM sigmooidales es mucho menor que la SENS en el caso de RFB-SVM. En resumen, aunque las RBF-SVM presentan mejores resultados que las S-SVM, ninguno de estos modelos aportan mejora a los resultados obtenidos mediante LDA.

Tabla IV.28. Resultados de los distintos modelos para la clasificación de pimentones dulces.

Modelo	Extremadura		Murcia		Global	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA	82 ± 14	56 ± 22	56 ± 22	82 ± 14	72 ± 10	66 ± 13
S-SVM	90 ± 14	17 ± 26	17 ± 26	90 ± 14	58 ± 8	49 ± 13
RFB-SVM	86 ± 16	36 ± 25	36 ± 25	85 ± 16	64 ± 12	58 ± 14
MLP-ANN 4:7:2	98 ± 5	86 ± 18	86 ± 18	98 ± 5	93 ± 8	91 ± 12

SENS: sensibilidad; SPEC: especificidad.

El empleo de ANN permite la construcción de modelos más eficaces en la diferenciación de tipo dulce según el origen. Se obtiene un modelo de arquitectura 4:7:2, utilizando las cuatro variables. En este caso se han empleado funciones de activación simples como lo son la logística e identidad. El número de neuronas dentro de la capa oculta suele determinarse por medio de ensayo error [45].

Los resultados que se presentan en la tabla IV.28. muestran una mejora con respecto a LDA y SVM. La SENS para las muestras de EX es 98% mientras que la SENS para el origen MU es 86%.

IV.3.1.2. Diferenciación geográfica de muestras de pimentones picantes a partir del parámetro ASTA y coordenadas CIELab.

Para los pimentones picantes se sigue el mismo esquema que en el caso anterior. Al realizarse el análisis PCA se comprueba que los dos primeros PCs son los que muestran autovalores mayores que la unidad, siendo estos 2,21 y 1,34 para PC1 y PC2, respectivamente. Con estos dos primeros componentes se explica el 88,7% de la varianza, ya que PC1 explica el 55,2% y PC2 el 33,5%. El siguiente PC, con un autovalor de 0,43 explica por sí mismo un 11,2% por lo que con los tres primeros se explica más de un 99% de la varianza total.

En la tabla IV.29.se muestran las correlaciones de las variables con los PCs. En este caso existe una fuerte correlación entre PC1 y las variables a* y b*. PC2 está más correlacionado con L* y ASTA.

Tabla IV.29. Pesos de las variables en los PCs

	PC1	PC2	PC3	PC4
ASTA	-0,54	0,71	0,46	0,00
L*	-0,25	-0,89	0,38	-0,03
a*	-0,96	0,05	-0,24	-0,09
b*	-0,97	-0,21	-0,11	0,10

Cuando se representa la distribución de las muestras en el plano formado por PC1 y PC2 en la figura IV.70. se observa que la tendencia de las muestras de EX es agruparse en la zona negativa de PC1 extendiéndose a lo largo de PC2, mientras que las muestras de MU lo hacen en torno a valores negativos de PC2, con valores positivos y negativos de PC1.

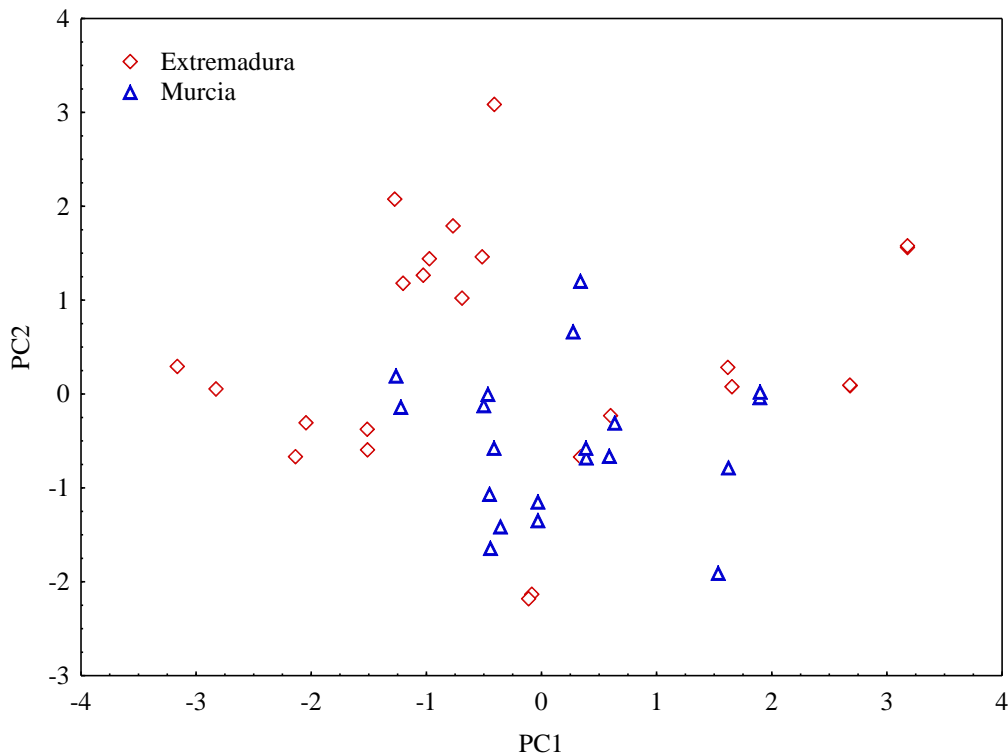


Figura IV.70. Distribución de las muestras de pimentón del tipo picante en el plano formado por PC1 y PC2.

Utilizando las cuatro variables elegidas para la diferenciación se crean dos modelos LDA, uno estándar considerando todas las variables y otro de inclusión progresiva, FLDA, donde las variables se van introduciendo según su poder discriminante. En este caso, el FLDA consideró como variables más discriminantes L* y b*, dejando el parámetro ASTA y a* fuera del modelo. En las siguientes figuras IV.71.

y IV.72., se representan los resultados del LDA estándar y de FLDA, respectivamente, observándose solapamiento entre muestras de los dos orígenes para los dos modelos considerados.

Los resultados obtenidos en ambos casos, presentados en la tabla IV.30., son muy similares. A la luz de estos resultados, se considera la posibilidad de mejorar la diferenciación y para ello se prueban técnicas capaces de resolver problemas no lineales.

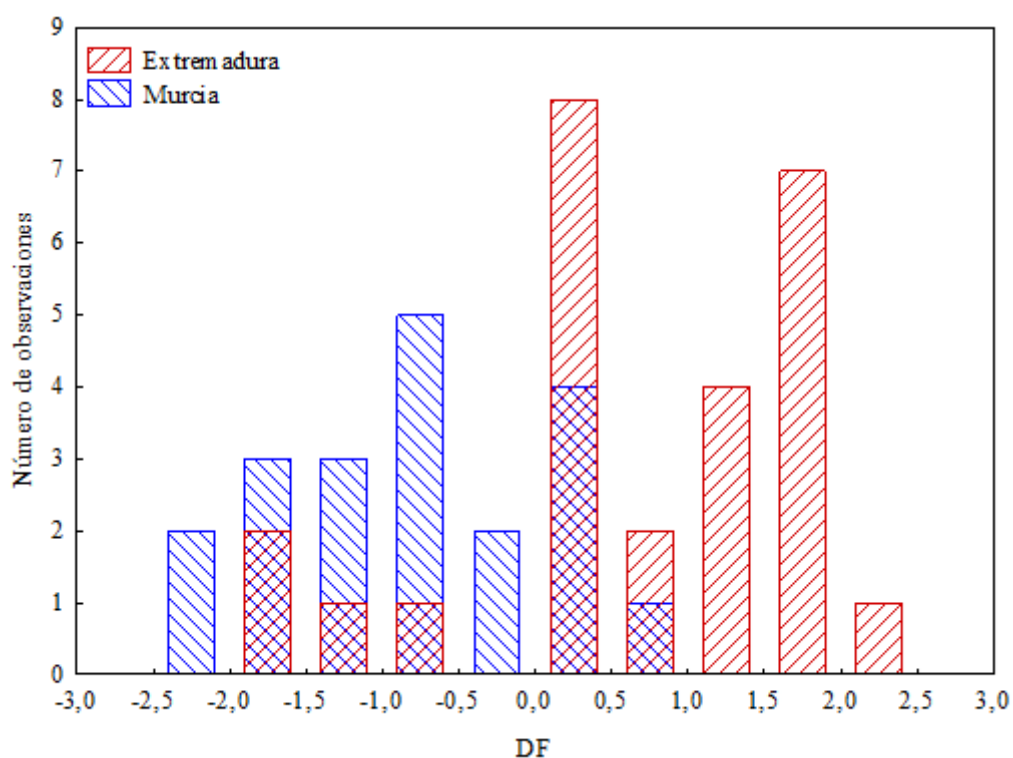


Figura IV.71. Distribución de las muestras de pimentones picante en DF según el origen, para LDA estándar.

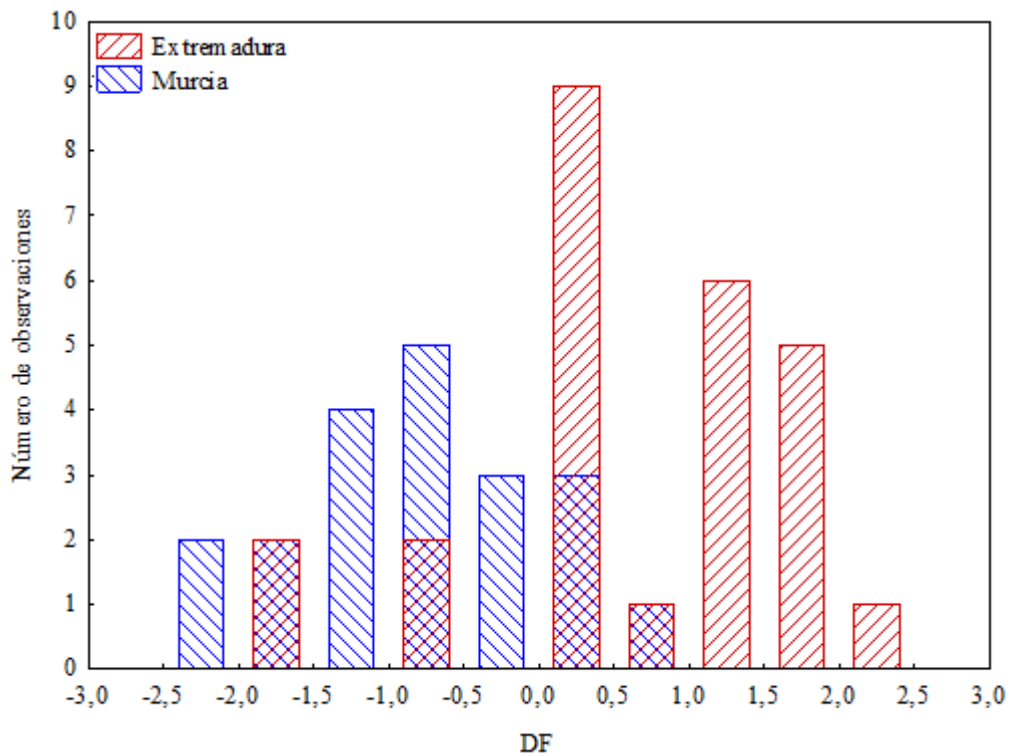


Figura IV.72. Distribución de las muestras de pimentones del tipo picante en DF según el origen para FLDA.

Se han obtenido cuatro modelos de clasificación SVM. Dos modelos usando las cuatro variables y cambiando el tipo de Kernel (RBF, y S). Por otra parte, se crean otros dos modelos empleando las variables señaladas por LDA, L^* y b^* , y los dos tipos de Kernel citados. Los resultados obtenidos se muestran en la tabla IV.30. Si se analizan los resultados de SENS y SPEC globales se observa que el empleo de las dos variables más discriminantes introduce una mejora en los resultados. Además, el Kernel sigmoidal ofrece mejores resultados ya sea empleando cuatro o dos variables. Las desviaciones estándar de estos modelos son altas, como en el caso del pimentón dulce, aunque sí es cierto que la SENS y la SPEC mejoran con respecto a los resultados de SVM en la clasificación del tipo dulce.

Para conseguir modelos más adecuados, se han usado ANNs multicapa para los dos casos posibles, uno empleando las cuatro variables de las que se parte, y otro empleando las dos variables más discriminantes de acuerdo al modelo FLDA. Las redes tienen la arquitectura 4:5:2 y 2:4:2, respectivamente. Se emplean en este caso funciones de activación del tipo tangencial y logística, y el número de capas ocultas se elige mediante ensayo error [46].

Tabla IV.30. Resultados de los distintos modelos para la clasificación de pimentones picantes.

Modelo	Extremadura		Murcia		Global	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA ^a	82 ± 10	68 ± 12	68 ± 12	82 ± 10	75 ± 7	73 ± 8
FLDA ^a	86 ± 5	68 ± 12	68 ± 12	86 ± 5	78 ± 5	75 ± 7
S-SVM ^a	78 ± 18	49 ± 33	49 ± 33	78 ± 18	65 ± 11	62 ± 15
RFB-SVM ^a	69 ± 17	71 ± 16	71 ± 16	69 ± 17	70 ± 11	70 ± 11
S-SVM ^b	81 ± 15	63 ± 30	63 ± 30	81 ± 15	73 ± 13	71 ± 16
RFB-SVM ^b	79 ± 12	70 ± 15	70 ± 15	79 ± 12	75 ± 9	74 ± 10
MLP-ANN ^a -4:5:2	98 ± 5	91 ± 10	91 ± 10	98 ± 5	94 ± 4	94 ± 5
MLP-ANN ^b -2:4:2	97 ± 6	90 ± 11	90 ± 11	97 ± 6	94 ± 7	93 ± 8

a: modelo calculado con las variables ASTA, L*, a* y b*. b: modelos construido con L* y b*. SENS: sensibilidad; SPEC: especificidad.

Los resultados obtenidos de SENS y SPEC, presentados en la tabla IV.30., son mejores cuando se aplica ANN que cuando se han empleado LDA o SVM. Los resultados globales están por encima del 90%, y si se analizan de forma individual encontramos que la SENS de la clase EX es del 97% y la de MU del 90%. Aunque en el caso de las SVMs los resultados mejoraban cuando se emplean las variables consideradas como más discriminantes según FLDA, en el caso de las ANN se obtienen mejores resultados cuando el modelo se construye a partir de las cuatro variables originales.

IV.3.1.3. Diferenciación geográfica de muestras de pimentones agridulces a partir del parámetro ASTA y coordenadas CIELab.

Cuando se lleva a cabo el PCA en el caso de las muestras de pimentón agridulce se observa que los PCs que presentan autovalores mayor que la unidad son PC1 y PC2, con valores de 2,15 y 1,68, respectivamente. En este caso los tres primeros PCs explican una varianza total del 99,6%. Si se consideran solo PC1 y PC2, la varianza acumulada que explican es de un 95,8%, de la cual PC1 explica un 53,8% y PC2 un 42%. La correlación de las variables con los componentes principales se muestra en la tabla IV.31. Las correlaciones más importantes son las que existen entre las variables ASTA y

A continuación, como primer paso para la obtención de un modelo capaz de diferenciar orígenes, aplicamos un LDA estándar a las cuatro variables, y otro FLDA. Este último modelo incluye las cuatro variables iniciales por lo que los resultados no varían con respecto al primer LDA estándar. No se obtiene una buena clasificación, como se puede apreciar en la figura IV.74. y en los resultados de la tabla IV.32. Los valores de SENS para la clase EX son mejores que para MU, obteniéndose unos resultados globales de SENS y SPEC están en torno al 70%.

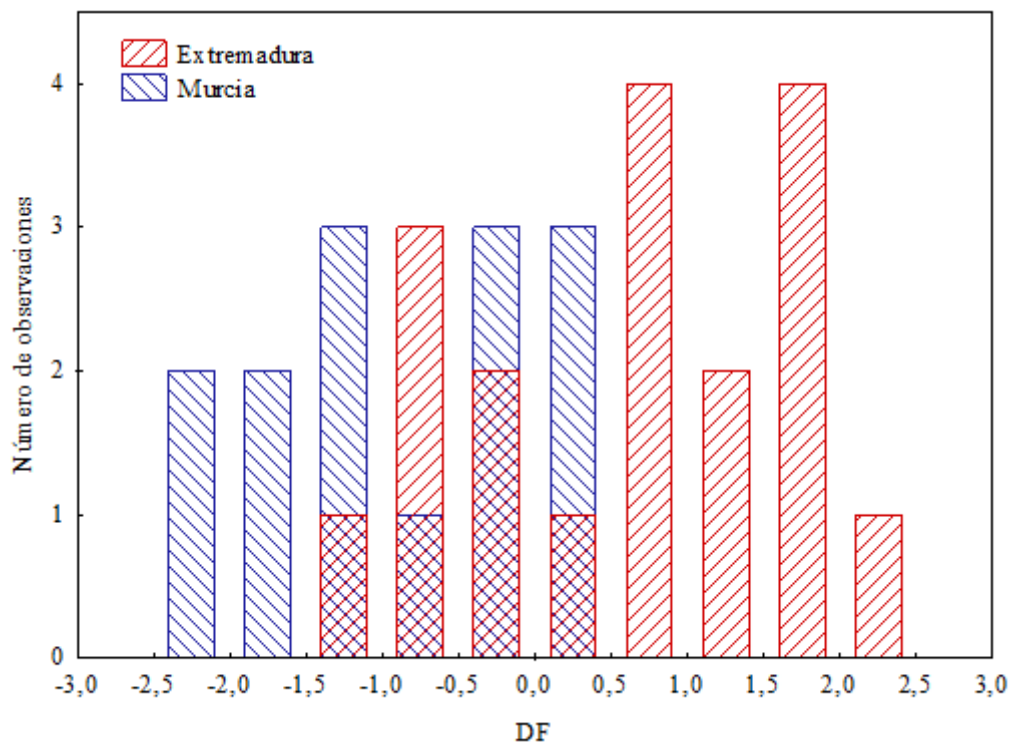


Figura IV.74. Distribución de las muestras de pimentones del tipo agridulce en DF según origen para LDA estándar.

Se procede a la construcción de modelos mediante SVM para comprobar si mejoran los resultados aplicando un método no lineal. Se han generado dos modelos empleando las cuatro variables, uno con Kernel sigmooidal, S-SVM, y otro con Kernel de función de base radial, RBF-SVM.

Con respecto a LDA; los resultados individuales de SENS de EX muestran una mejoría. Sin embargo, tanto para la SENS de MU, como globalmente, no se produce ninguna mejora. Además, las desviaciones estándar son muy elevadas cuando se analiza

la clasificación por tipos, lo que nos da una idea de que el modelo no es robusto. Los resultados, que se muestran en la tabla IV.32, indican que los valores individuales de SENS y SPEC son mejores cuando se emplea RFB-SVM que al aplicar S-SVM.

Tabla IV.32. Resultados de los distintos modelos para la clasificación de pimentones agridulces.

Modelo	Extremadura		Murcia		Global	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA	80 ± 14	62 ± 16	62 ± 16	80 ± 14	72 ± 11	70 ± 11
S-SVM	93 ± 12	11 ± 18	11 ± 18	93 ± 12	56 ± 3	48 ± 5
RFB-SVM	85 ± 16	40 ± 25	40 ± 25	85 ± 16	64 ± 9	61 ± 10
MLP-ANN-4:3:2	98 ± 6	92 ± 14	92 ± 14	98 ± 6	95 ± 8	95 ± 9

Se construye un modelo de ANN multicapa, MLP-ANN, empleando las cuatro variables y cuya arquitectura es 4:3:2 que mejora notablemente los resultados. Se obtienen valores de SENS y SPEC globales en torno al 95% en ambos casos. La SENS a la hora de clasificar las muestras de MU es mayor que en el caso de EX. Con las ANN se consigue una separación que no se había conseguido con otros métodos de reconocimiento de patrones lineales y no lineales.

IV.3.2. Estudio del color en base al espectro ultravioleta visible.

En los siguientes apartados se pretende diferenciar el origen de los pimentones dulces, picantes y agridulces a partir de los datos de absorbancia obtenidos tras registrar los espectros en el rango de 250 a 800 nm. Siguiendo con el esquema de la figura IV.62. una vez registrado el espectro, se elige el rango de longitudes de onda de 380 a 780 nm, y las absorbancias correspondientes a esas longitudes de onda son utilizadas como variables de entrada para los modelos que se van a construir. Al ser un número elevado de variables y además presentar una elevada correlación, se emplea PCA para disminuir el número de variables y generar unas nuevas no correlacionadas entre sí. Los valores de cada muestra para esas nuevas variables, denominados scores, se emplean para construir los modelos de clasificación LDA, SVM o ANN.

IV.3.2.1. Diferenciación geográfica de muestras de pimentones dulces a partir del espectro UV-vis.

Se registran los espectros de las muestras de pimentón de los dos orígenes en el rango de longitudes de onda de 380 a 780 nm. Para llevar a cabo el estudio quimiométrico se genera una matriz de datos con 400 columnas, una por cada valor de absorbancia a las longitudes de onda desde 380 a 780nm, y 34 filas, una por cada muestra analizada (14 muestras de Murcia y 20 de Extremadura). Se aplica PCA para reducir el número de variables y observar posibles tendencias. En la figura IV.75.se muestra el espectro de absorción promedio (línea continua) \pm la desviación estándar de las muestras de pimentón de los dos orígenes considerados (línea discontinua).

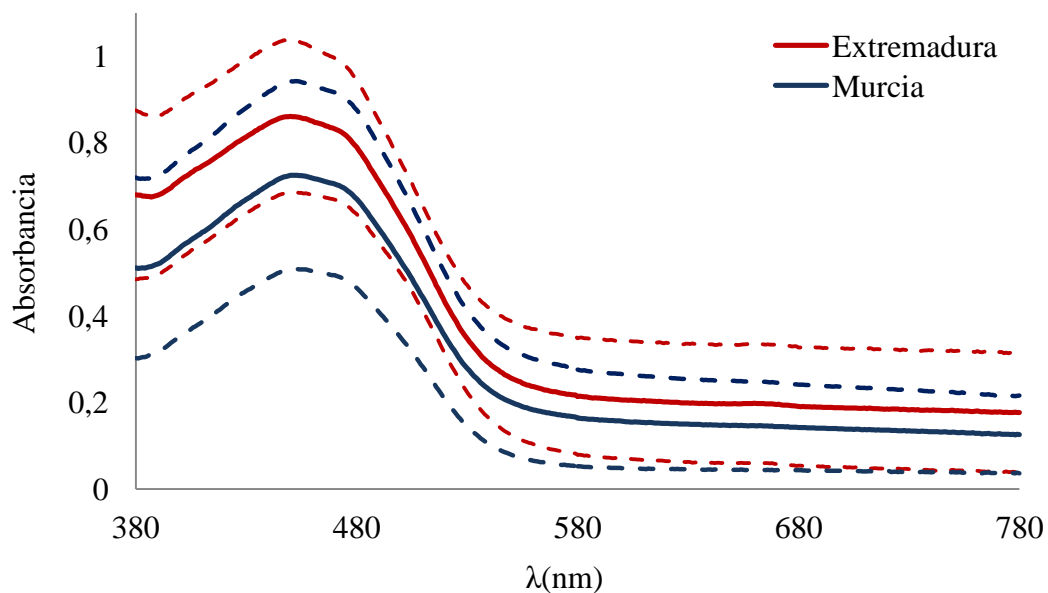


Figura IV.75. Espectro medio (línea continua) \pm desviación estándar (línea discontinua) de las muestras de pimentón dulce de Extremadura y Murcia.

De acuerdo al criterio de Kaiser [27], el número de PCs necesarios para explicar la tendencia de los datos, viene marcado por aquellos cuyo autovalor es superior a la unidad. En este caso pretendemos construir modelos de clasificación a partir de los valores obtenidos para las muestras en el espacio de los PCs. Aunque los 4 primeros PCs cumplen el criterio de Kaiser, se retiene la información de un quinto PC para una eventual utilización en los posteriores modelos de clasificación.

En la figura IV.76.se pueden observar las correlaciones de los PCs con las absorbancias a cada longitud de onda. El PC1 aparece correlacionado con longitudes de onda a partir de 490 nm, mientras que PC2 está más correlacionado con la primera parte del espectro. Pesos superiores a 0,7 indican una fuerte correlación, mientras que el intervalo 0,7-0,4 indica una correlación moderada. Valores menores que 0,4 muestran que no existe correlación entre las variables y los PCs [28]. En este caso, PC3, PC4 y PC5 no presentan correlaciones importantes, ya que en ningún caso toman valores superiores a 0,7.

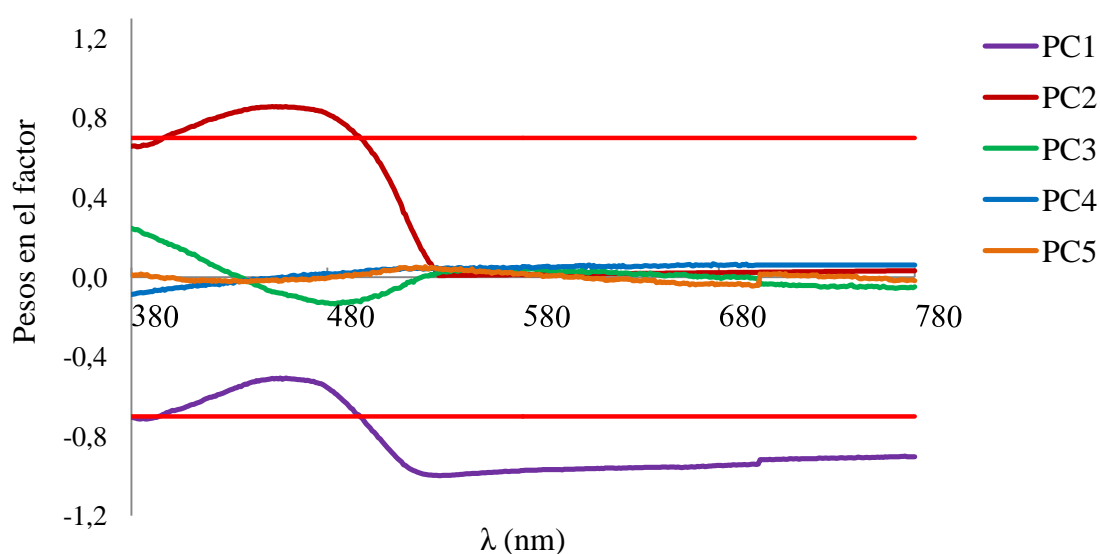


Figura IV.76. Correlación de los PCs con las absorbancias a las longitudes de onda indicadas para las muestras de pimentón dulce.

Los dos primeros PCs son los que mayor porcentaje de varianza explican. El PC1 explica un 73,2%, mientras que PC2 un 25,7%. Con los cuatro primeros PCs, de autovalores 293,47, 103,23, 2,27 y 1,69, el porcentaje de varianza explicado es del 99,9%.

Cuando se representan las muestras en el plano de PC1 y PC2, figura IV.77., se puede observar que no hay una separación clara en función de su origen. Las muestras de Murcia están generalmente agrupadas en torno a valores positivos de PC1 y positivos y negativos de PC2, aunque varias de las muestras se agrupan a valores negativos de ambos PCs. En el caso de las de Extremadura, las muestras aparecen dispersas en los cuatro cuadrantes, siendo imposible describir una tendencia.

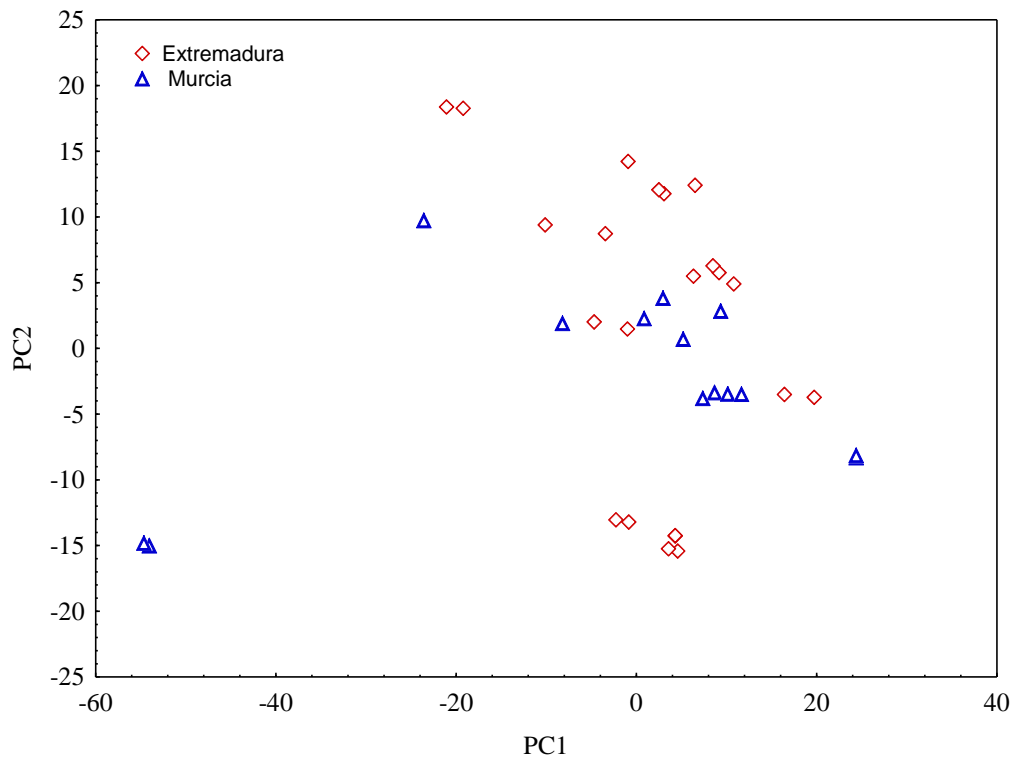


Figura IV.77. Distribución de las muestras de pimentón del tipo dulces en el plano de PC1 y PC2.

Una de las técnicas empleadas para construir modelos de clasificación es LDA. El primer modelo de LDA se construye empleando los PCs con autovalores superiores a la unidad, es decir, los cuatro primeros PCs. Los resultados que se obtienen se muestran en la tabla IV.33. El valor de SENS en el caso de las muestras de EX es mayor que para las muestras de MU. Las desviaciones estándar son elevadas, por lo que la robustez del método no se puede asegurar. El hecho de que no presenten una clara separación entre orígenes, aunque sí cierta tendencia a separarse, puede observarse en la figura IV.78., donde se representan las muestras de EX y MU según su coordenada para la función discriminante. Si se realiza un análisis de inclusión progresiva partiendo de los cuatro primeros PCs, nos encontramos con que el modelo queda definido por muy pocas variables sin mejora de los resultados (datos no incluidos).

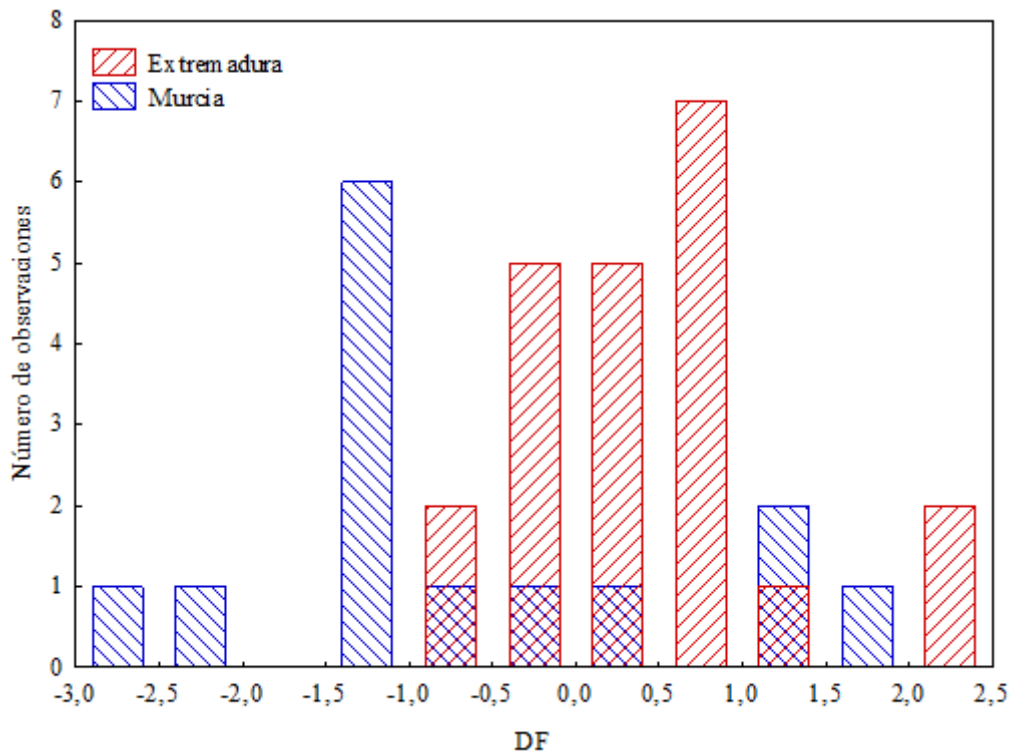


Figura IV.78. Distribución de las muestras de pimentón del tipo dulce en DF según el origen para LDA con 4PCs.

El segundo modelo LDA que se realiza emplea los cinco primeros PCs. El gráfico de la figura IV.79. muestra que la separación tampoco es completa, aunque los resultados de la tabla IV.33. indican que existe una cierta mejora en los valores, globales y por clases, de SENS y SPEC. Analizando los resultados globales, cuando se incluye una variable más al modelo, se tienen resultados de SENS y SPEC del 85 y 80%, respectivamente, mientras que cuando se calcula el modelo a partir de los cuatro primeros PCs se tienen resultados que no alcanzan el 80%.

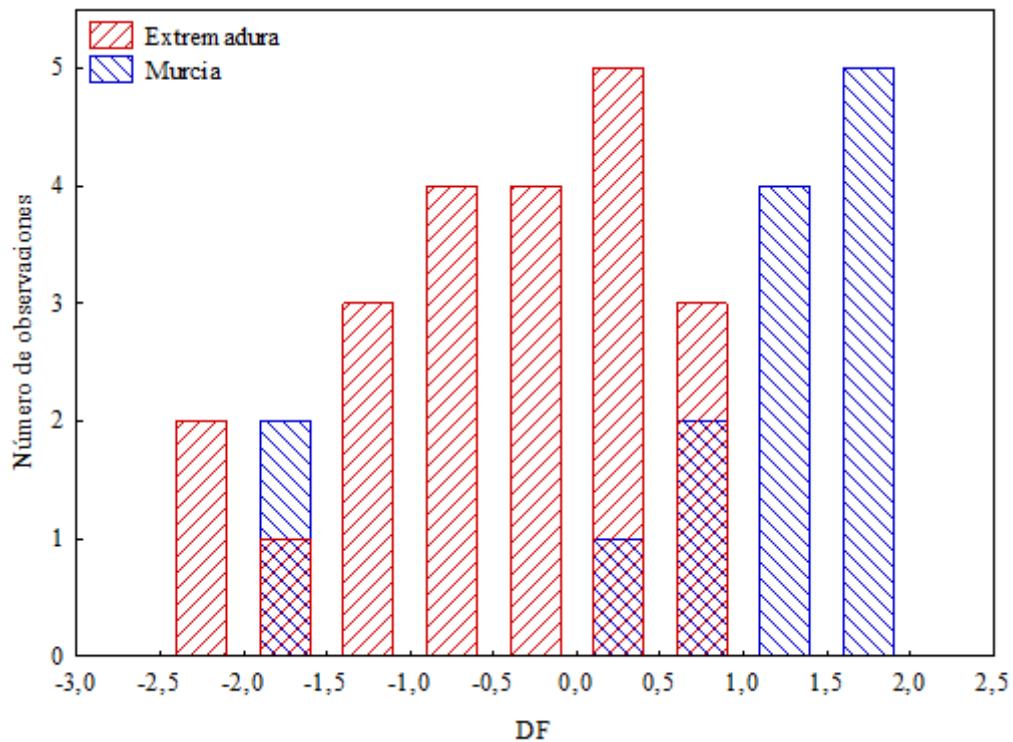


Figura IV.79. Distribución de las muestras de pimentón del tipo dulce en DF según el origen para LDA con 5PCs.

Con el fin de mejorar los resultados obtenidos por LDA, se han realizado modelos mediante SVM. Se aplican dos tipos de funciones Kernel [49], S y RFB, dando lugar a dos modelos distintos.

Igual que en el caso de LDA se construyen modelos con 4 y 5 PCs. Los resultados de los cuatro modelos se calculan a partir de nueve replicados, realizando la validación cruzada DAGJK, y los parámetros SENS y SPEC se presentan en la tabla IV.33.

Los resultados globales logrados mediante SVM cuando se emplean 4 PCs no mejoran los que se obtenían con los modelos LDA. Analizados de forma individual, los resultados de SENS para clasificar las muestras de Extremadura son mucho mejores cuando se emplean SVM que LDA, sin embargo, la SENS de Murcia empeora con respecto al empleo de LDA. En el caso de las S-SVM con 5 PCs, la SENS de MU es mucho mayor que la de EX pero los resultados globales siguen sin introducir ninguna

mejoría. Además las desviaciones estándar de los replicados de validación cruzada vuelven a presentar valores muy altos, con lo que no se obtienen modelos robustos.

Tabla IV.33. Resultados de los distintos modelos para la clasificación de pimentones dulces.

Modelo	Extremadura		Murcia		Global	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA ^a	82 ± 11	64 ± 22	64 ± 22	82 ± 11	75 ± 12	71 ± 15
LDA ^b	88 ± 9	80 ± 10	80 ± 10	88 ± 9	85 ± 9	83 ± 9
S-SVM ^a	96 ± 9	33 ± 26	33 ± 27	96 ± 9	72 ± 10	57 ± 16
RFB-SVM ^a	93 ± 9	52 ± 20	52 ± 20	93 ± 9	78 ± 8	68 ± 12
S-SVM ^b	33 ± 26	96 ± 9	96 ± 9	33 ± 26	72 ± 10	57 ± 16
RFB-SVM ^b	88 ± 14	53 ± 22	53 ± 22	88 ± 14	74 ± 12	66 ± 15
MLP-ANN ^a -4:7:2	93 ± 7	91 ± 15	91 ± 15	93 ± 7	92 ± 8	92 ± 10
MLP-ANN ^b -5:8:2	96 ± 9	98 ± 7	98 ± 7	96 ± 9	97 ± 7	97 ± 6

a: modelo calculado con los cuatro primeros PCs. b: modelos construido con los cinco primeros PCs.
SENS: sensibilidad; SPEC: especificidad.

Se construyen modelos mediante MLP-ANN con 4 y 5 PCs, con la subdivisión necesaria en set de entrenamiento, verificación y test. La red construida con los cuatro primeros PCs tiene una arquitectura 4:7:2, mientras que la segunda red construida a partir de los 5 primeros PCs presenta 5 neuronas en la capa de entrada, 8 en la oculta y 2 en la capa de salida. Los resultados de ambos modelos se recogen en la tabla IV.33. En este caso sí se mejoran los resultados de SENS y SPEC cuando se obtiene el modelo MLP-ANN con un PC de más con respecto a los elegidos según el criterio de Kaiser. Los valores de SENS y SPEC están en torno al 92% para el primer modelo ANN y al 97% para el segundo. Por tanto, es posible clasificar y diferenciar pimentones del tipo dulce según el origen mediante datos del espectro UV-vis y MLP-ANN.

IV. 3.2.2. Diferenciación geográfica de muestras de pimentones picantes a partir del espectro UV-vis.

Se han registrado los espectros de las muestras de pimentones del tipo picante y distintos orígenes en el rango de longitud de onda de 250 a 800 nm. En la figura IV.79 se representan los valores medios de las longitudes de onda de las 48 muestras medidas en el rango de 380-780 nm, rango empleado en el estudio de diferenciación. Las líneas

discontinuas representan los valores promedios de las absorbancias más y menos las desviaciones estándar correspondientes. En este caso se emplean 20 muestras del origen Murcia (MU) y 28 de Extremadura (EX).

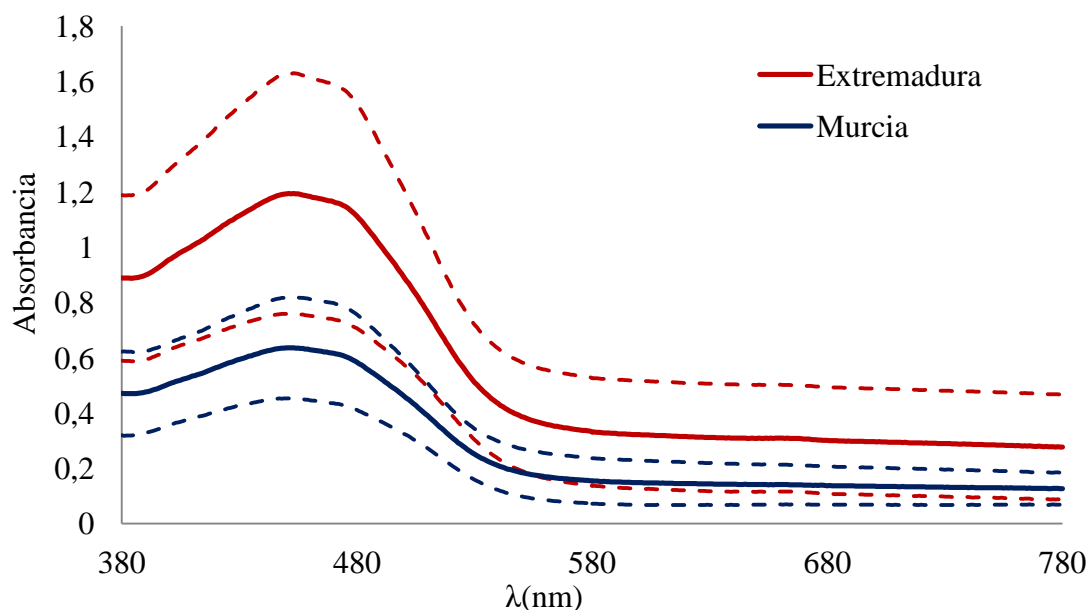


Figura IV.80. Espectro medio (línea continua) \pm desviación estándar (línea discontinua) de las muestras de pimentón picante de Extremadura y Murcia.

Al igual que se hizo en el caso de los pimentones dulces, se genera la matriz de datos con 400 columnas, una por cada longitud de onda desde 380 a 780nm, y 48 filas una por cada muestra analizada. Con el PCA se reduce el número de variables. Se utiliza el criterio de Kaiser para elegir las variables retenidas, PC1 y PC2 en este caso, aunque se ha retenido la información de PC3 para comprobar la posibilidad de alguna mejora tras su inclusión en los modelos de clasificación.

En la figura IV.81. se observa nuevamente cómo las correlaciones más importantes son las que presentan PC1 a partir de 480 nm y PC2 en la primera zona del espectro. PC3 no presenta correlaciones por encima de 0,7 en ninguna zona del espectro.

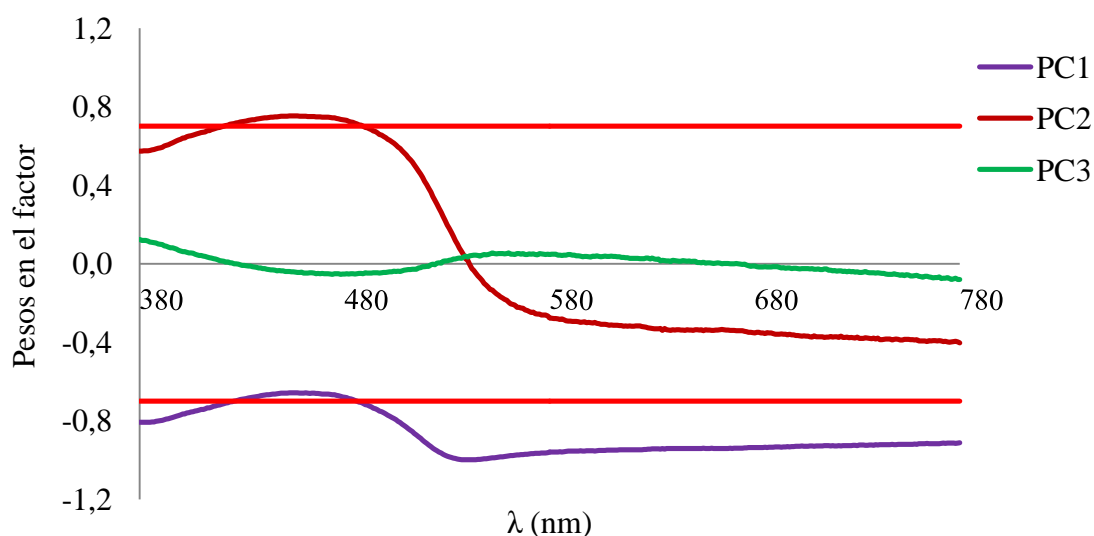


Figura IV.81. Correlación de los PCs con las absorbancias a las longitudes de onda indicadas para las muestras de pimentón picante.

En cuanto a la varianza acumulada, PC1 y PC2 explican un 99,73% de la varianza total, presentando autovalores 308,14 y 91,78, respectivamente. El tercer PC, con un autovalor de 0,79, explica un 0,20% de la varianza total.

Al representar las muestras en el espacio definido por los dos primeros PCs se obtiene la figura IV.81., donde no se aprecia una separación clara entre clases, pero sí se observa cierta tendencia por parte de las muestras de EX y MU a distribuirse de manera diferente en el plano. Las muestras de EX se agrupan en su mayoría en torno a valores positivos de PC2 y negativos de PC1, y las muestras de MU presentan valores positivos de PC1 y negativos de PC2.

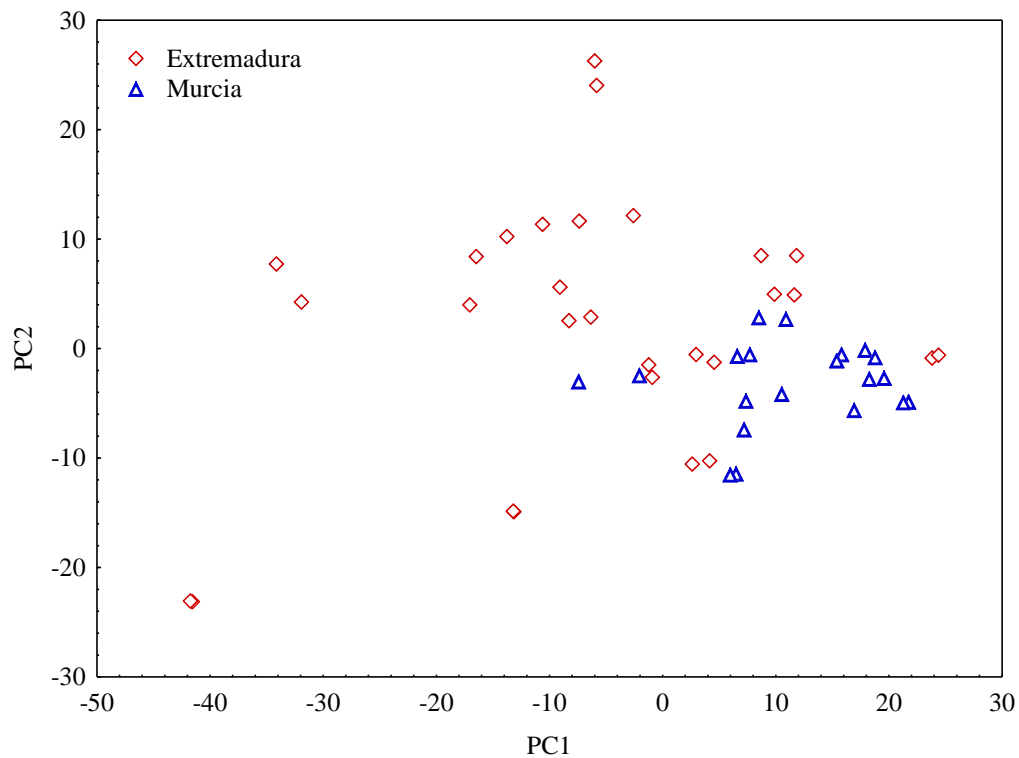


Figura IV.82. Distribución de las muestras de pimentones picantes en el plano de PC1 y PC2.

Se han construido dos modelos de clasificación mediante LDA estándar, uno empleando los dos primeros PCs y otro con los tres primeros.

En el primer caso, no se obtiene una diferenciación completa de las muestras según origen (figura IV.83.). Los resultados, que se muestran en la tabla IV.34., son del orden de los obtenidos en los apartados anteriores donde la SENS y SPEC se mueven en el rango de 70-85% aproximadamente.

Cuando se emplean los tres primeros PCs para el cálculo del modelo LDA estándar, los resultados empeoran ligeramente con respecto a los obtenidos para dos PCs. En la figura IV.84. se observa cómo existe solapamiento entre muestras de los dos orígenes cuando se representan según la DF del modelo.

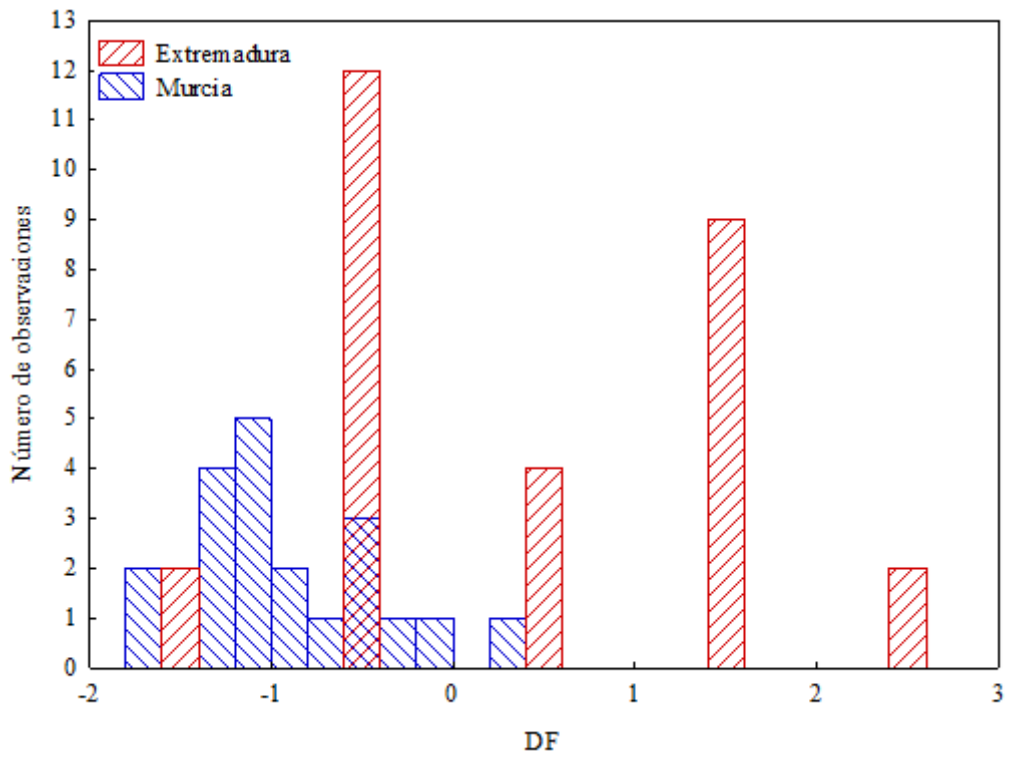


Figura IV.83. Distribución de las muestras de pimentón del tipo picante en DF según el origen para LDA con 2PCs.

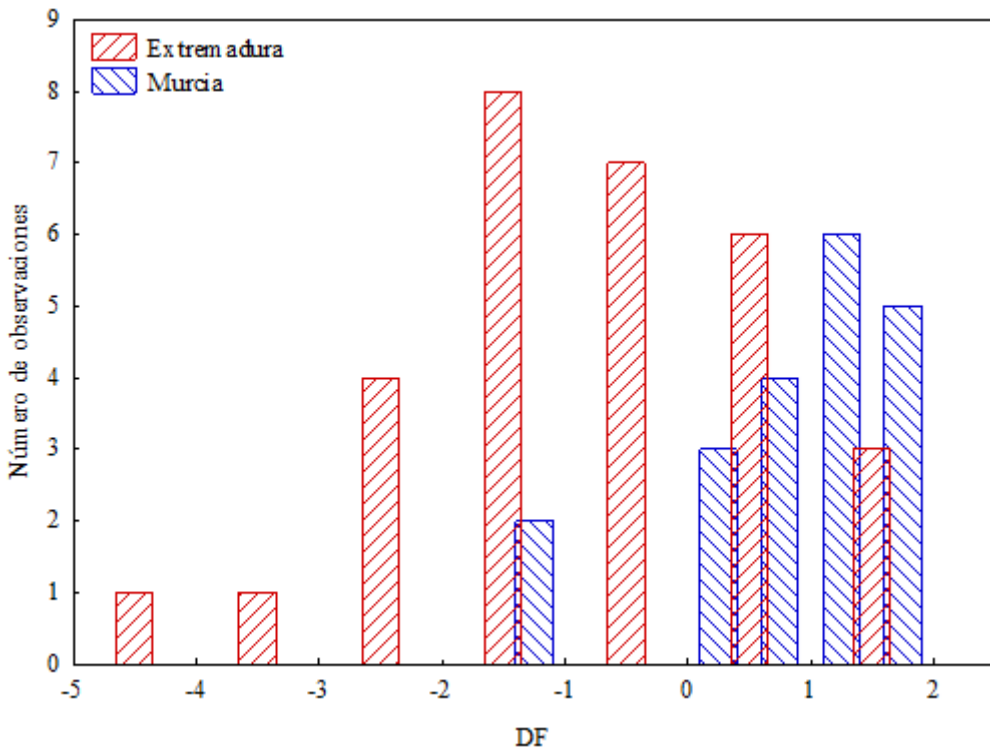


Figura IV.84. Distribución de las muestras de pimentón del tipo picante en DF según el origen para LDA con 3PCs.

Se crean cuatro modelos SVM combinando el tipo de Kernel, RFB o S, con el uso de 2 o 3 PCs. Analizando los resultados presentados en la tabla IV.34., se puede observar como cuando se realiza la clasificación mediante SVM empleando los tres primeros PCs, los resultados empeoran con respecto a los dos modelos propuestos LDA. Sin embargo, los resultados fruto de emplear el modelo RFB-SVM con dos PCs son comparables a los obtenidos mediante LDA.

Tabla IV.34. Resultados de los distintos modelos para la clasificación de pimentones picantes.

Modelo	Extremadura		Murcia		Global	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA ^a	84 ± 9	89 ± 14	89 ± 14	84 ± 9	86 ± 9	87 ± 10
LDA ^b	79 ± 6	87 ± 13	87 ± 13	79 ± 6	82 ± 8	84 ± 10
S-SVM ^a	70 ± 17	79 ± 31	79 ± 31	70 ± 17	74 ± 10	76 ± 14
RFB-SVM ^a	77 ± 11	90 ± 10	90 ± 10	77 ± 11	82 ± 7	85 ± 6
S-SVM ^b	74 ± 14	74 ± 29	74 ± 29	74 ± 14	74 ± 7	75 ± 13
RFB-SVM ^b	76 ± 17	76 ± 31	76 ± 31	76 ± 17	76 ± 10	76 ± 15
MLP-ANN ^a -2:6:2	94 ± 5	92 ± 10	92 ± 10	94 ± 5	94 ± 6	93 ± 7
MLP-ANN ^b -3:6:2	94 ± 5	93 ± 11	93 ± 11	94 ± 5	94 ± 3	94 ± 5

a: modelo calculado con los dos primeros PCs. b: modelos construido con los tres primeros PCs. SENS: sensibilidad; SPEC: especificidad.

Para intentar mejorar los resultados obtenidos con LDA y SVM se han construido modelos de clasificación MLP-ANN. Se crean dos modelos, MLP-ANN-2:6:2 y MLP-ANN-3:6:2, para 2 PCs y 3 PCs, respectivamente. La ANN basada en el uso de 3 PCs, presenta mejores resultados de clasificación que la de arquitectura 2:6:2. En la tabla IV.34. se recogen estos resultados en base a la SENS y SPEC, con resultados del 94 % para ambos parámetros.

IV.3.2.3. Diferenciación geográfica de muestras de pimentones agridulces a partir del espectro UV-vis.

En el caso del tipo agridulce, se tienen 18 muestras de Extremadura (EX) y 14 muestras correspondientes al origen Murcia (MU). En la figura IV.85. se representan mediante líneas continuas el espectro de absorción promedio y las líneas discontinuas representan los límites al sumar y restar la desviación estándar de dicha medidas para cada origen.

Se genera una matriz de datos con 400 columnas, una por cada longitud de onda desde 380 a 780nm, y 32 filas, una por cada muestra analizada.

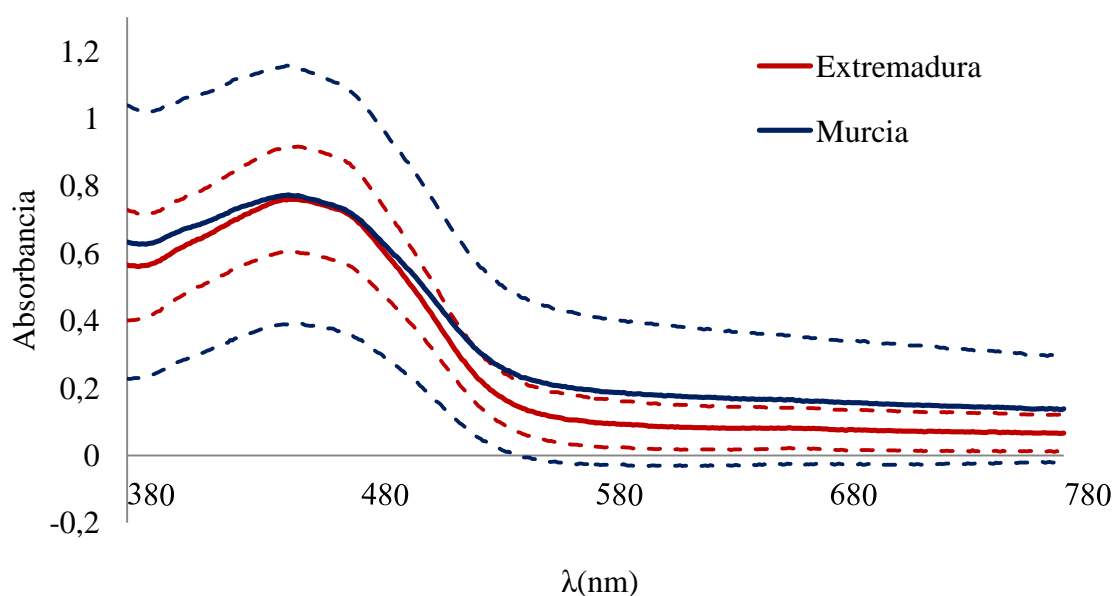


Figura IV.85. Espectro medio (línea continua) \pm desviación estándar (línea discontinua) de las muestras de pimentón agridulce de Extremadura y Murcia.

Cuando se realiza el PCA, se obtiene que los tres primeros PCs son los que presentan autovalores mayores a la unidad (373,43, 22,69 y 1,5), y al igual que en los casos anteriores, se retiene la información de un PC más.

Analizando estos datos se observa que un 99,9% de la varianza total se retiene con los tres primeros PCs. Los pesos de cada variable en los distintos PCs se muestran en la figura IV.86., donde se observa que el PC1 está muy correlacionado con todo el

espectro y el PC2 está más influenciado por las absorbancias en el rango de 440 a 490 nm, si bien los coeficientes de correlación son menores que 0,7 en este caso.

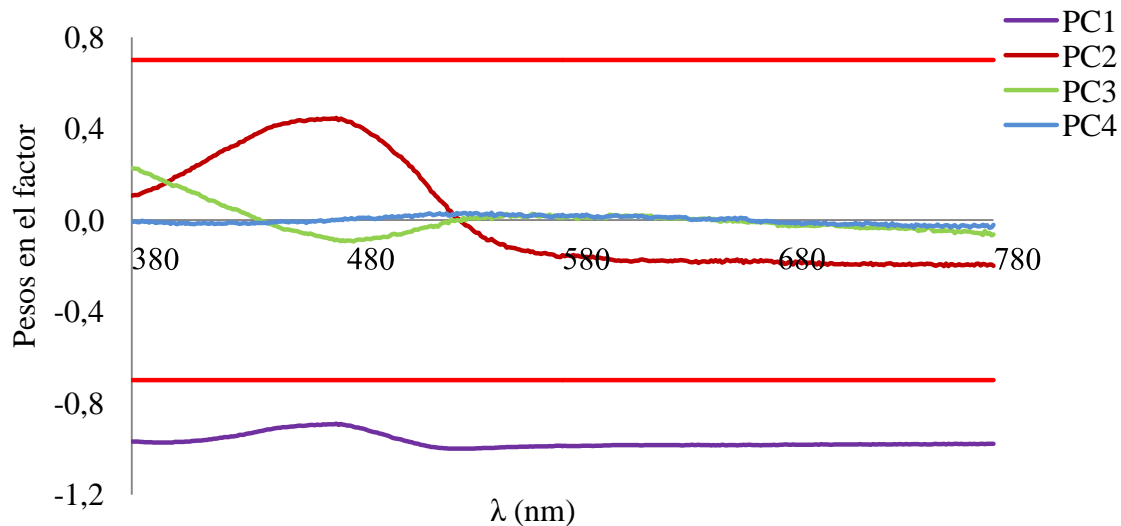


Figura IV.86. Correlación de los PCs con las absorbancias a las longitudes de onda indicadas para las muestras de pimentón agridulce.

Cuando se representan los scores de las muestras en el plano PC1 - PC2 (figura IV.87.) se observa como las muestras de Extremadura se agrupan mayoritariamente a valores positivos de PC2 mientras que las de Murcia lo hacen a valores negativos. Las muestras de Murcia se presentan a valores positivos de PC1 y las de Extremadura se distribuyen en torno a valores positivos y negativos de PC1. Puesto que se observa cierta tendencia natural a la separación de las muestras de los distintos orígenes, se recurre a los modelos de clasificación para conseguir esta separación a partir de las nuevas variables generadas tras el análisis PCA.

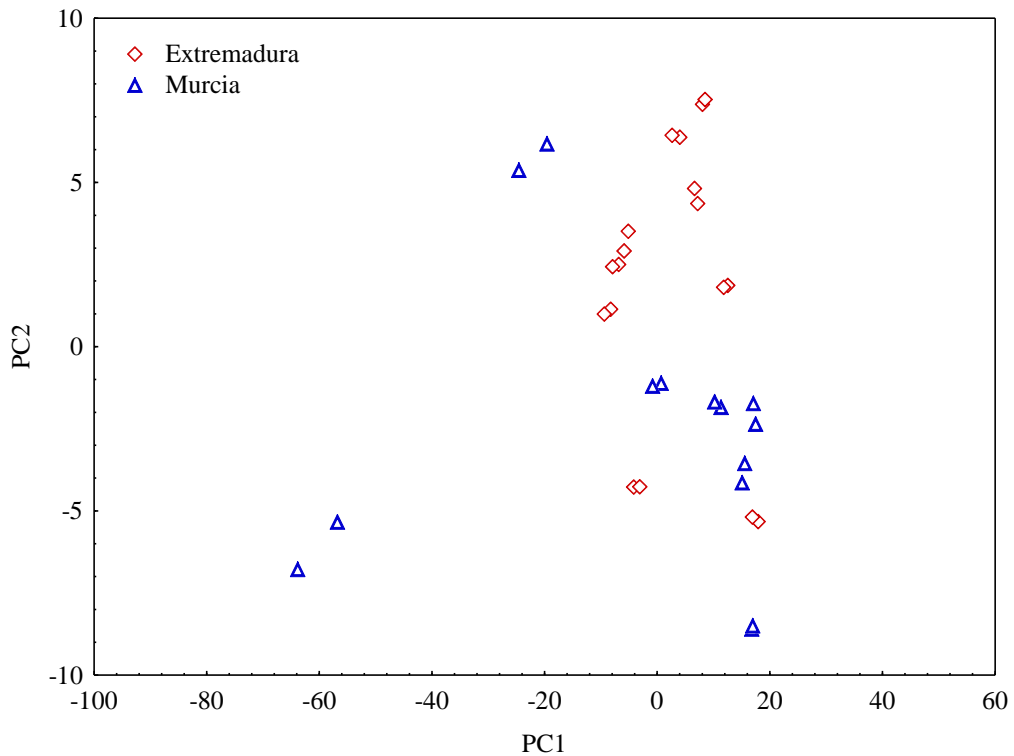


Figura IV.87. Distribución de las muestras del tipo agridulce en el plano de PC1 y PC2.

Los primeros modelos de clasificación se construyen mediante LDA. En primer lugar se obtiene un modelo LDA empleando los PCs con autovalores superiores a la unidad, es decir, los tres primeros. En la figura IV.88. se observa la distribución de las muestras según la función discriminante calculada. Como se puede observar, el modelo no diferencia adecuadamente las muestras. Tras validar el modelo mediante DAGJK se obtienen los valores de SENS y SPEC que se muestran en la tabla IV.35.

Los valores de SENS son mayores para el caso de las muestras de Extremadura, siendo éstos del 80% en este caso mientras que para las muestras de Murcia encontramos valores del 76%. Las desviaciones estándar toman, en ambos casos, valores muy altos por lo que el método no puede considerarse robusto.

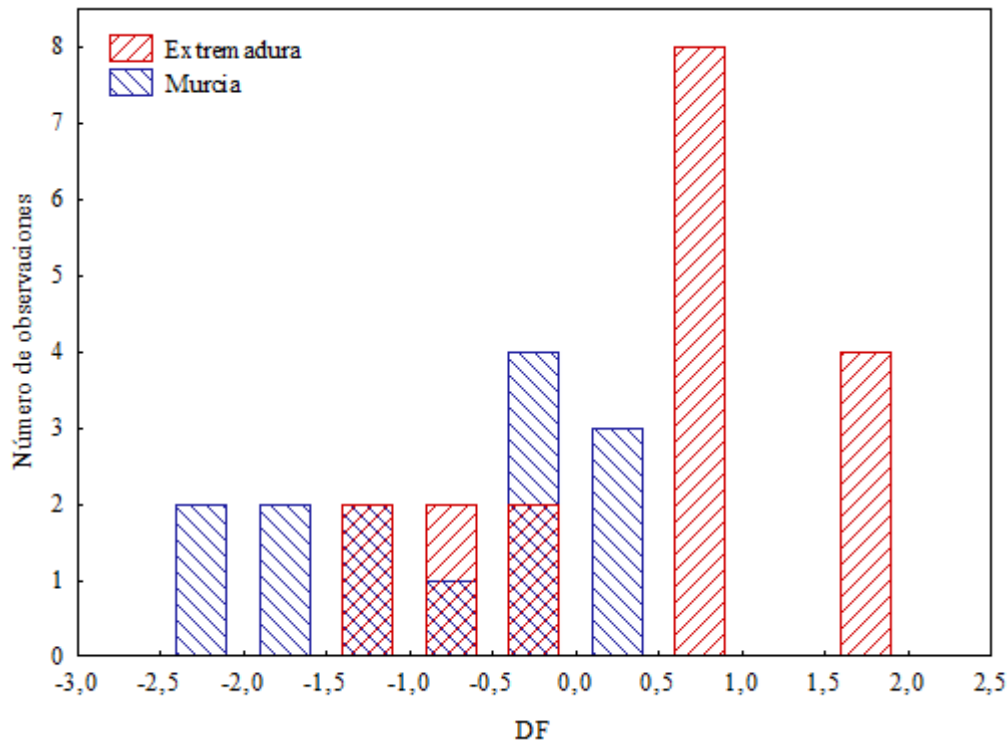


Figura IV.88. Distribución de las muestras de pimentón del tipo agridulce en DF según el origen para LDA con 3PCs.

Para comprobar si estos resultados son mejorables introduciendo información adicional del sistema, se construye un nuevo modelo LDA usando el cuarto PC. En la figura IV.89. se observa que el modelo no separa adecuadamente las muestras, pero en los datos de la tabla IV.35. se observa cierta mejoría.

En este caso, los valores de SENS para las muestras de Extremadura siguen siendo mayores que los de Murcia, con valores del 92% y 75%, respectivamente. Las desviaciones estándar siguen siendo altas, al igual que en el modelo LDA^a (3PCs).

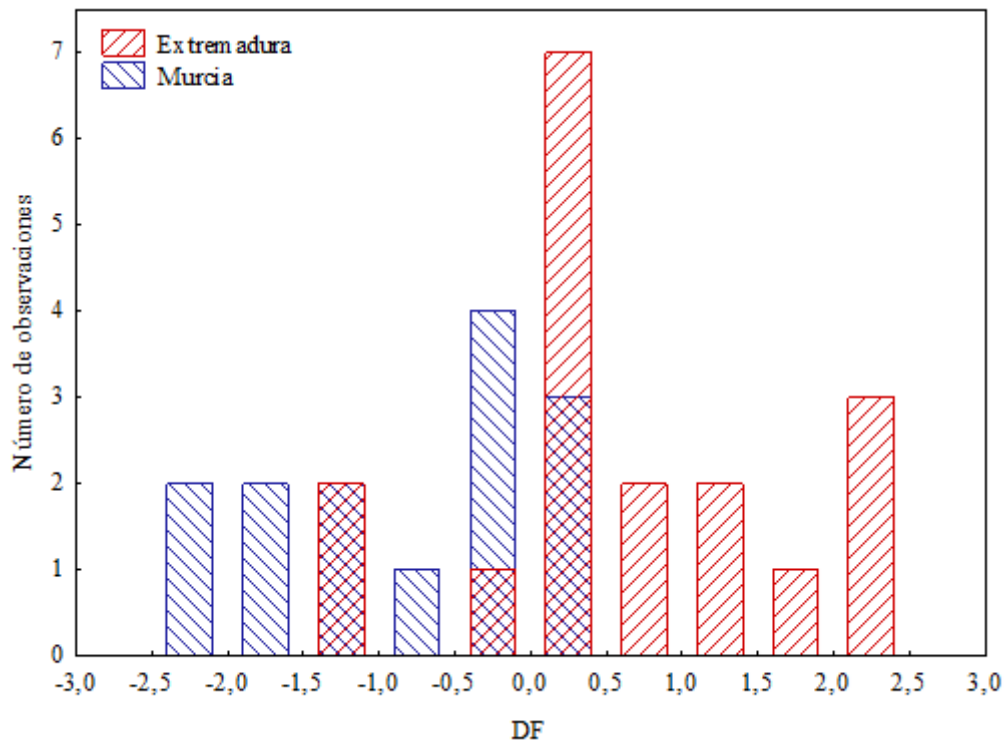


Figura IV.89. Distribución de las muestras de pimentón del tipo agridulce en DF según el origen para LDA con 4PCs.

Con el objetivo de mejorar estos resultados, se recurre a SVM. Se aplican dos tipos, S-SVM y RFB-SVM, dando lugar a cuatro modelos distintos. En ambos casos se partió de los tres primeros PCs, y posteriormente se usaron los cuatro primeros PCs. Se realizó la validación cruzada mediante DAGJK. Los resultados de ambos modelos en términos de SENS y SPEC se muestran en la tabla IV.35.

Tabla IV.35. Resultados de los distintos modelos para la clasificación de pimentones agridulces.

Modelo	Extremadura		Murcia		Global	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA ^a	80 ± 16	76 ± 12	76 ± 12	80 ± 16	78 ± 7	78 ± 5
LDA ^b	92 ± 10	75 ± 14	75 ± 14	92 ± 10	84 ± 6	83 ± 7
S-SVM ^a	85 ± 21	28 ± 31	28 ± 31	85 ± 21	59 ± 11	54 ± 13
RFB-SVM ^a	84 ± 21	30 ± 34	30 ± 34	84 ± 21	58 ± 13	55 ± 14
S-SVM ^b	72 ± 35	42 ± 39	42 ± 39	72 ± 35	55 ± 14	60 ± 10
RFB-SVM ^b	78 ± 34	41 ± 41	41 ± 41	78 ± 34	57 ± 16	61 ± 11
MLP-ANN ^a -3:9:2	93 ± 7	90 ± 12	90 ± 12	96 ± 7	94 ± 5	93 ± 6
MLP-ANN ^a -4:6:2	89 ± 14	92 ± 12	92 ± 13	89 ± 14	90 ± 8	91 ± 8

a: modelo calculado con los tres primeros PCs. b: modelos construido con los cuatro primeros PCs. SENS: sensibilidad; SPEC: especificidad.

Las SVM no incluyen mejoras en los resultados globales con respecto a LDA. En el caso de la SENS para el origen Extremadura cuando se calcula el modelo con 3 PCs, se observa que el empleo de ambos SVM proporciona resultados mejores. Si se analiza de forma individual la SENS en la clasificación de las muestras de MU ocurre lo contrario, disminuyendo del 70% cuando se aplica LDA al 30% si se clasifican las muestras según modelos SVMs.

Par mejorar estos resultados se emplean modelos mediante redes neuronales artificiales de perceptrones multicapa. En este caso se han creado dos modelos, MLP-ANN 3-9-2 y MLP-ANN 4-6-2, empleando 3 PCs y 4 PCs, respectivamente. Los resultados se muestran en la tabla IV.35. Para el modelo MLP-ANN 3-9-2 se obtienen valores globales de SENS y SPEC del 94% y 93% respectivamente. En el caso de MLP-ANN 4-6-2, estos valores son del 90% y 91% para SENS y SPEC, respectivamente.

A la luz de los resultados se puede concluir que el modelo que proporciona mejores resultados a la hora de clasificar pimentones del tipo agridulce en función de su origen es el basado en MLP-ANN. En este caso no existe mucha diferencia en cuanto a usar los tres primeros PCs o los cuatro primeros.

IV.3.3. Comparación de modelos de clasificación a partir de datos de parámetros ASTA y coordenadas CIELab y UV-vis.

En este apartado se van a comparar de manera gráfica los resultados de los distintos modelos, tanto los calculados a partir de los datos ASTA, L^* , a^* y b^* como los obtenidos a partir de los PC-scores de los datos de las absorbancias, para los distintos tipos de pimentón. Se presentan diagramas de barras donde aparece el porcentaje de SENS y SPEC obtenidos, mostrando la desviación estándar como una barra de error sobre la barra del porcentaje.

En la figura IV.90. se comparan los resultados obtenidos tras aplicar LDA, SVM y MLP-ANN a los datos de ASTA, L^* , a^* y b^* , en el caso se los pimentones dulces y se observa que el modelo que mejores resultados ha proporcionado es el MLP-ANN 4:7:2, con resultados cercanos al 100% para la SENS y la SPEC. Los resultados, tanto de SENS como de SPEC, obtenidos para LDA son mejores que los que proporcionan los dos tipos de SVM.

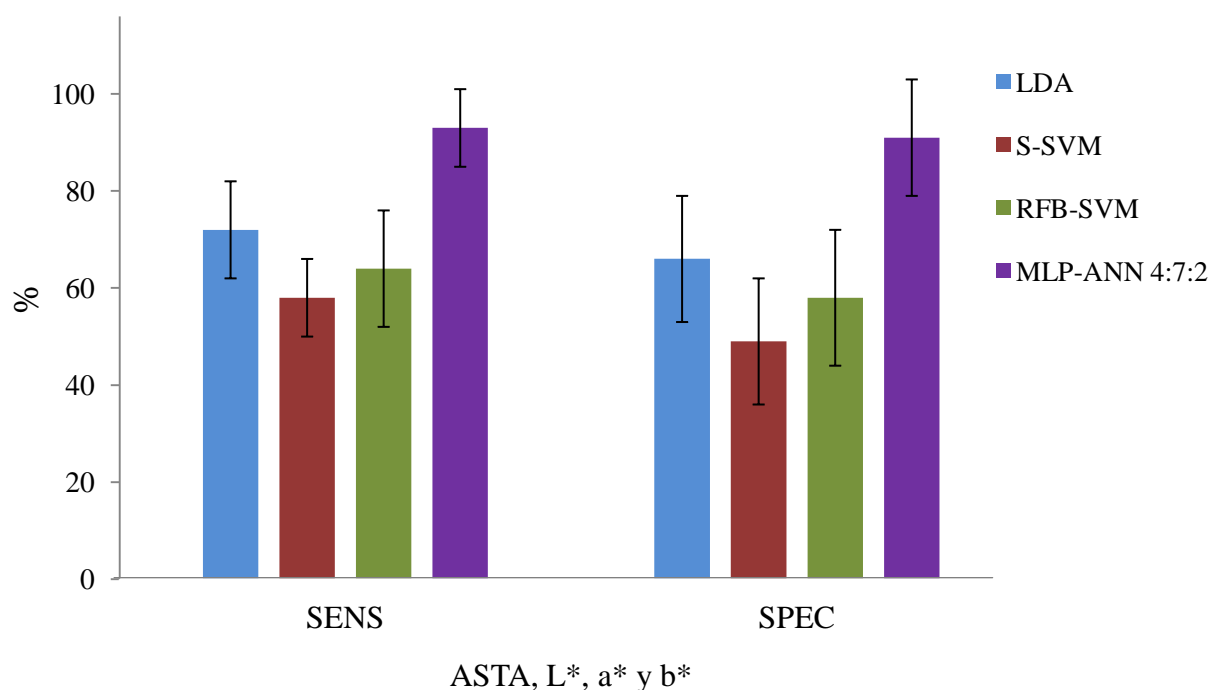


Figura IV.90. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón dulce a partir de ASTA, L^* , a^* y b^* .

En la figura IV.91. se representa el diagrama equivalente al anterior pero con los modelos obtenidos a partir de los PC-scores. En este caso, los modelos que proveen mejores resultados son los MLP-ANN, en concreto el MLP-ANN^b, que además presentan valores de desviación estándar menores. Si se compran los modelos LDA, se tiene que LDA^b proporciona mejores resultados. En el caso de RFB-SVM, el modelo RFB-SVM^a ofrece mayor porcentaje de SENS y SPEC. Ambos modelos de S-SVM presentan resultados similares y las desviaciones estándar semejantes. Eligiendo las mejores variantes de cada modelo, es decir, LDA^b, S-SVM^a, RFB-SVM^a y MLP-ANN^b, se pueden comparar con los resultados obtenidos cuando se calculan los modelos con el parámetro ASTA y las coordenadas CIELab. En la figura IV.92. se representan estos resultados.

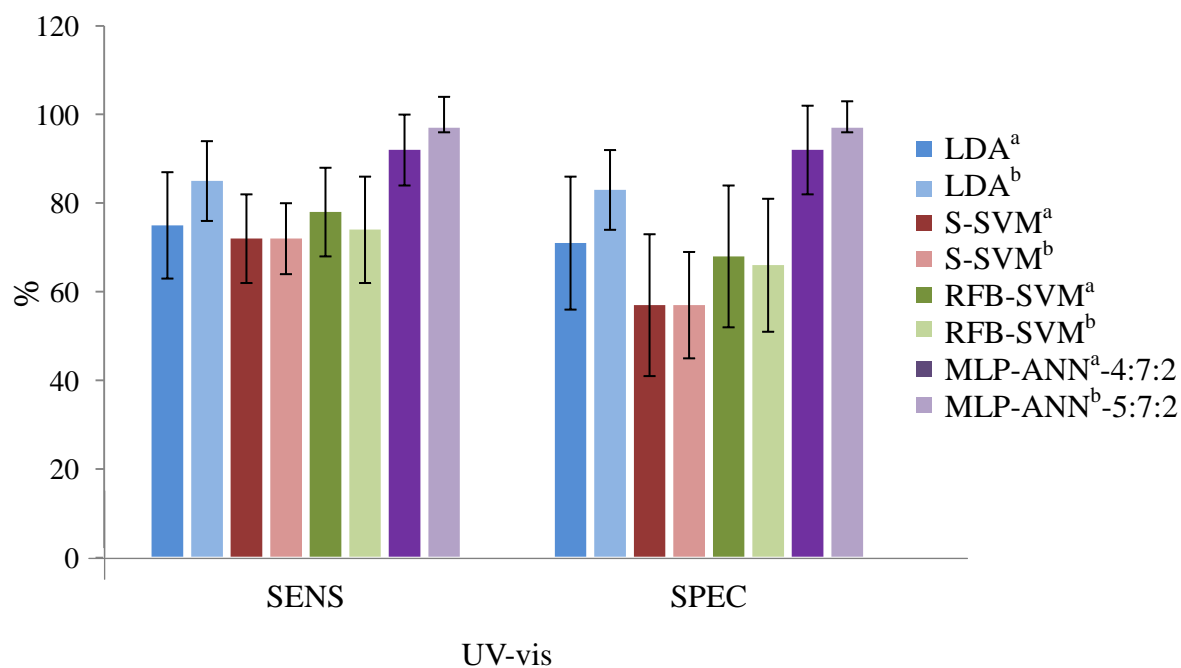


Figura IV.91. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón dulce a partir de UV-Vis.

Como se puede observar en la figura IV.92. cuando se comparan los modelos obtenidos mediante parámetros del color y mediante PC-scores, los mejores resultados en todos los casos son para los segundos, siendo el mejor de todos el modelo calculado mediante MLP-ANN.

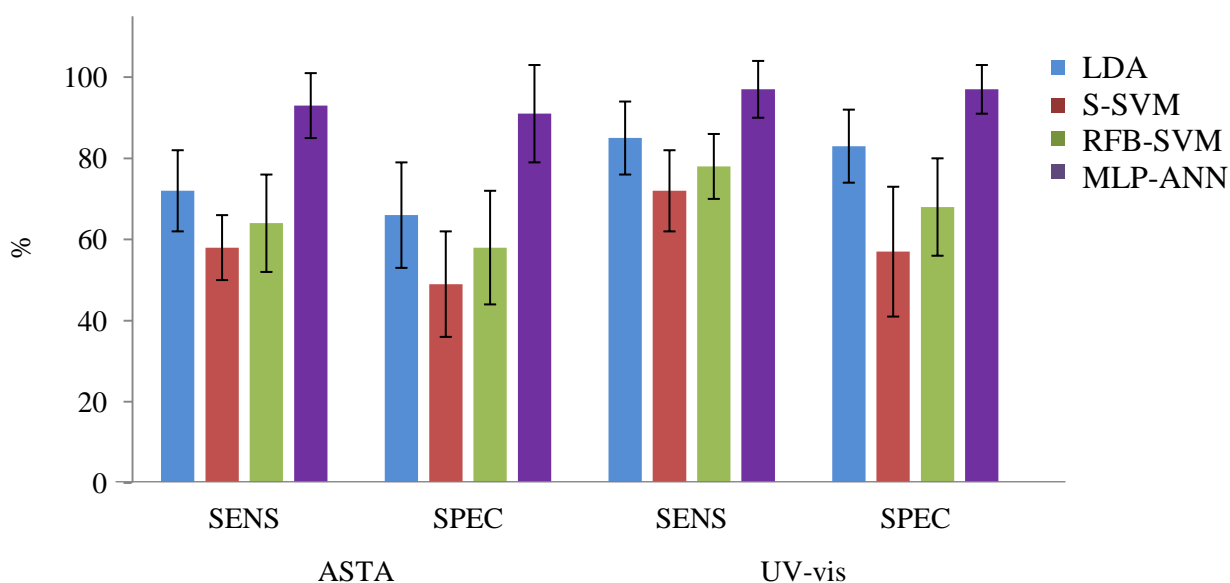


Figura IV.92. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón dulce a partir de parámetros del color (izquierda) y a partir de PC-scores (derecha).

En el caso de los pimentones de tipo picante, en la figura IV.93. se han comparado los resultados obtenidos tras aplicar LDA, SVM y MLP-ANN a los datos de ASTA, L^* , a^* y b^* . Se puede observar como el modelo que presenta mejores resultados es MLP-ANN-4:5:2. Los porcentajes de ambos modelos MLP-ANN son similares pero la desviación estándar de SENS y SPEC es menor en el caso en el que se usa el parámetro ASTA y las coordenadas L^* , a^* y b^* como variables de entrada. El resto de modelos proporciona mejores resultados cuando se usan las cuatro variables de entrada, excepto en el caso de LDA, que es mejor cuando sólo se emplean L^* y b^* . En este caso los modelos LDA son mejores que los SVM.

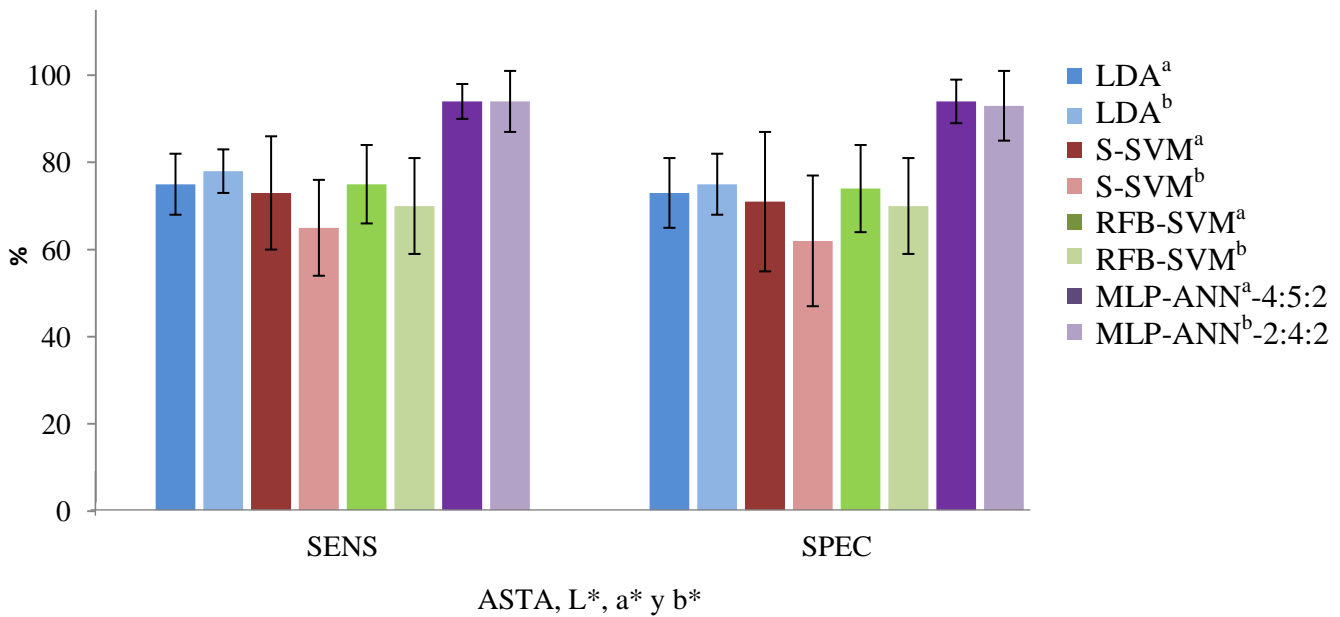


Figura IV.93. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón dulce a partir de ASTA, L*, a* y b*.

Si se realiza la comparación entre los modelos calculados con los valores de los PC-scores se observa, en la figura IV.94., que el modelo MLP-ANN empleando como variables de entrada los tres primeros PCs proporciona mejores resultados que el resto de modelos y valores de desviación estándar menores. El resto de modelos mejoran sus resultados cuando se emplean sólo los dos primeros PCs.

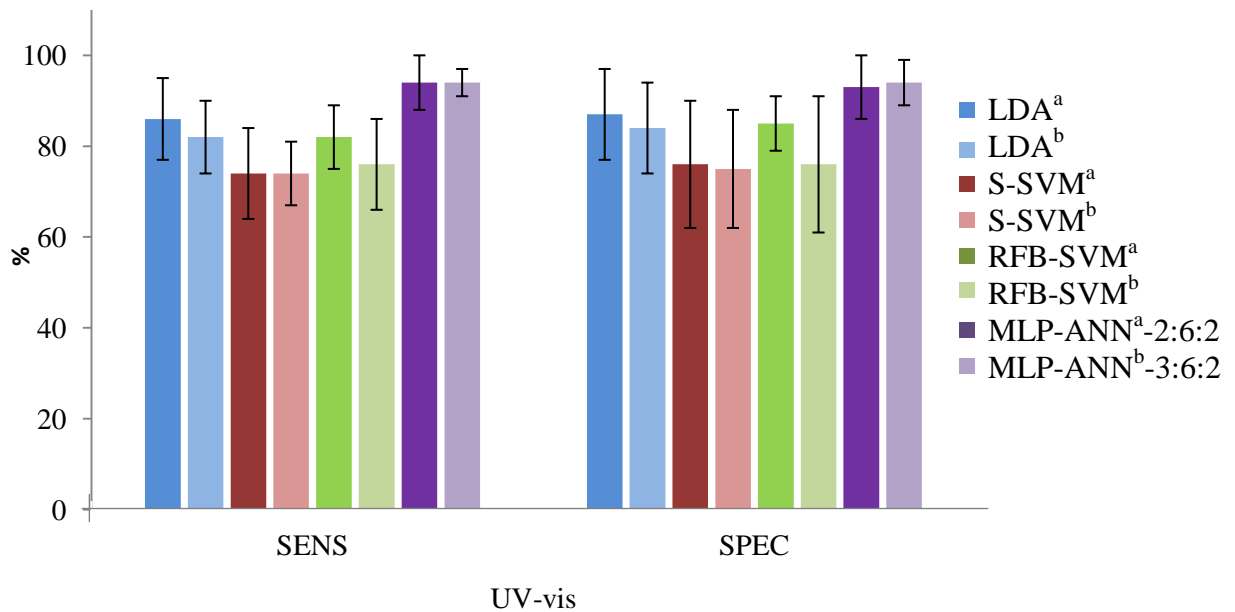


Figura IV.94. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón dulce a partir de los PC-scores.

Al igual que en el caso de los pimentones picantes, se eligen las mejores variantes de cada modelo para comparar los resultados obtenidos al utilizar los parámetros colorimétricos o los datos transformados mediante PCA. Los resultados obtenidos a partir de los PC-scores son generalmente mejores que los obtenidos a partir de los parámetros del color, excepto en el caso de RFB-SVM. En la figura IV.95., se observa como las MLP-ANN son los modelos con los que mejores resultados se obtienen, independientemente de la transformación de las variables de entrada empleadas. Por otra parte, los resultados obtenidos con MLP-ANN para la clasificación de pimentones dulces, son mejores que cuando hablamos de pimentones picantes, al igual que ocurría cuando se realizaba la clasificación según el perfil metálico.

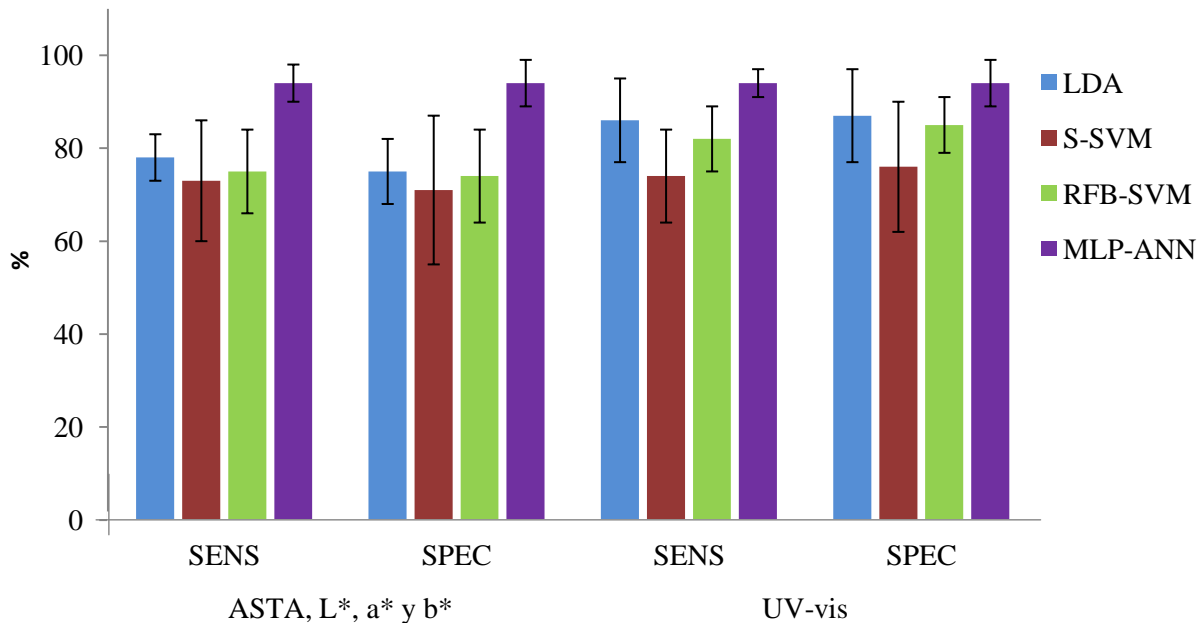


Figura IV.95. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón dulce a partir de los parámetros del color (izquierda) y de los PC-scores (derecha).

En el caso de los pimentones agridulces, las redes MLP-ANN vuelven a ser la mejor opción para clasificar muestras de ambos orígenes a partir de datos de ASTA, L*, a* y b*. El análisis LDA proporciona mejores resultados que cualquiera de los tipos de SVM, aunque las desviaciones estándar son mayores. Al igual que ocurría en los casos anteriores, las RFB-SVM ofrecen mejores resultados que las S-SVM. En la figura IV.96. se observan estos resultados, en forma de gráficos de barras con las desviaciones estándar.

Cuando se realiza la comparación entre los modelos que se han calculado a partir de los datos espectro UV-Vis transformados mediante PCA, se tiene que el modelo MLP-ANN con los tres primeros PCs proporciona mejores resultados que MLP-ANN con 4 PCs y las desviaciones estándar tanto de SENS como de SPEC son menores. En la figura IV.97. se observan estos hechos.

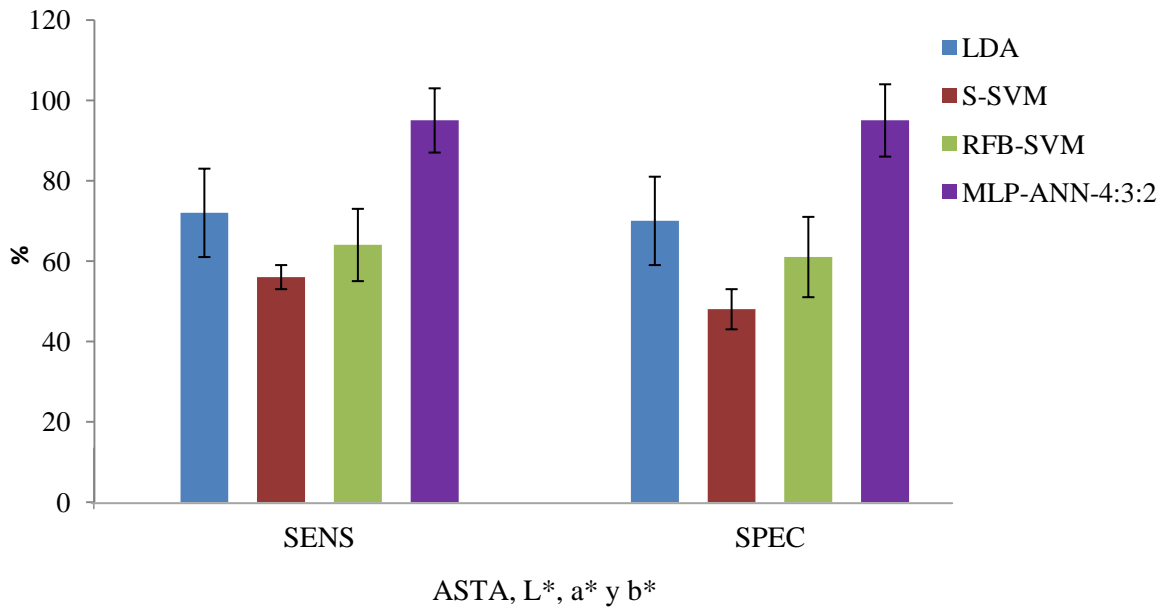


Figura IV.96. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón agrdulce a partir de ASTA, L*, a* y b*.

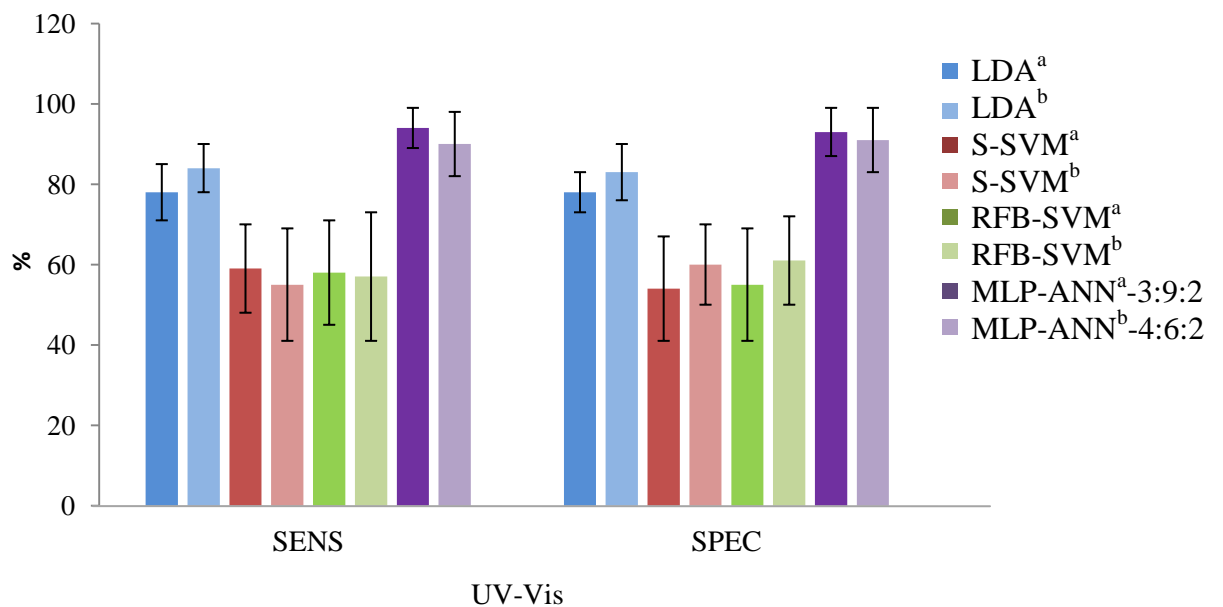


Figura IV.97. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón agrdulce a partir de los PC-scores.

Cuando se emplea LDA los resultados mejoran cuando se utilizan los 4 primeros PCs. En el caso de los dos tipos de SVM, la SENS aumenta cuando se usan 3PCs. Sin embargo la SPEC aumenta cuando se usan 4PCs.

Por último, se eligen las variantes de mejor resultado para cada modelo, el que emplea los parámetros del color y el que utiliza los PC-scores, con fines comparativos. Como se observa en la figura IV.98, los modelos MLP-ANN ofrecen mejores resultados con independencia de que se usen PC-scores o parámetros del color.

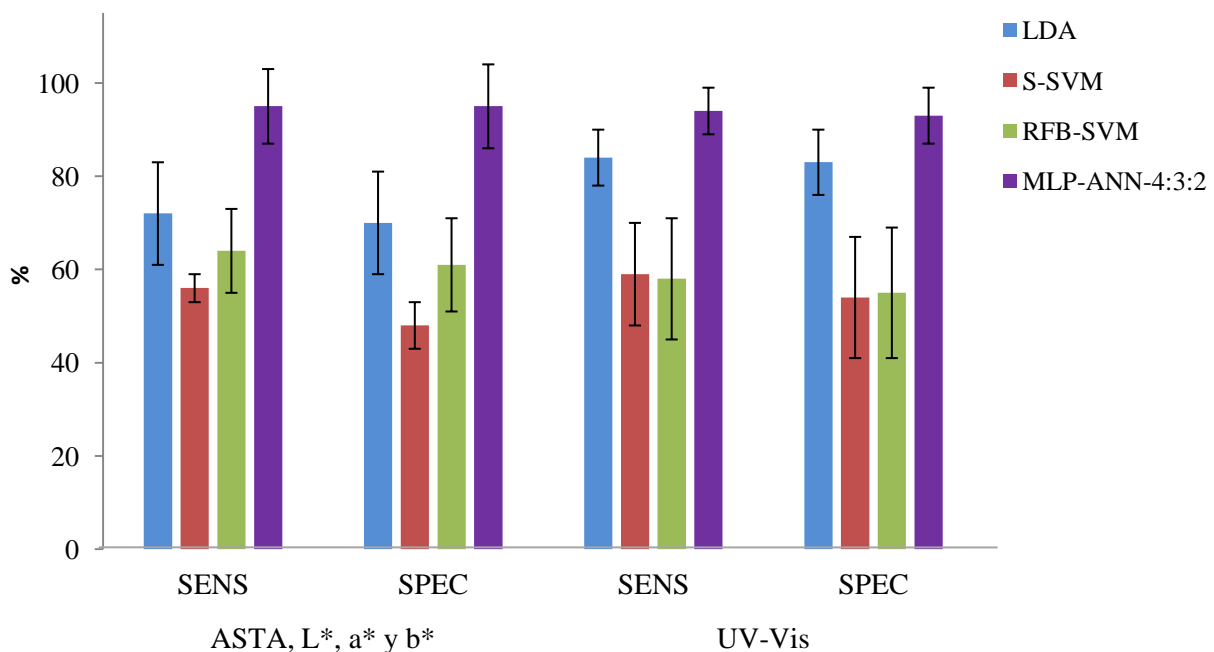


Figura IV.98. Diagrama de barras para la comparación de los modelos empleados en la clasificación geográfica de pimentón agridulce a partir de parámetros del color (izquierda) o PC-scores (derecha).

En el caso de los modelos construidos a partir de PC-scores los resultados de SENS y SPEC de LDA son mayores que cuando se usan las coordenadas colorimétricas y el parámetro ASTA. Las RFB-SVM son una mejor alternativa cuando se construyen a partir de ASTA y coordenadas CIELab, y las S-SVM cuando el modelo usa como variables de entrada los PCs correspondientes al estudio PCA del espectro UV-Vis.

IV.4. Determinación de capsaicinoides en pimentón mediante cromatografía líquida de alta eficacia

Los capsaicinoides son los responsables de la sensación picante de muchos frutos del género *Capsicum*, entre ellos los pimientos. En la bibliografía hay constancia de cinco capsaicinoides naturales: capsaicina (C), dihidrocapsaicina (DHC), nordihidrocapsaicina (NDHC), homocapsaicina (HC) y homodihidrocapsaicina (HDHC), cuyas estructuras se basan en amidas ácidas de vanilamina y cadenas C₉-C₁₁ de ácidos grasos [50]. En la figura IV.99. se representan las estructuras de los cinco capsaicinoides. La C y la DHC son los capsaicinoides principales en los pimientos, y por tanto en los pimentones, constituyendo alrededor del 90% de los capsaicinoides totales [51]. Por ello, son los que se van a considerar en este estudio. Los métodos más habituales para la determinación de C y DHC en muestras de pimiento y pimentón son GC y HPLC, aunque también se ha utilizado electroforesis capilar [52]. La técnica más habitual es HPLC, que es la que se va a utilizar en este trabajo.

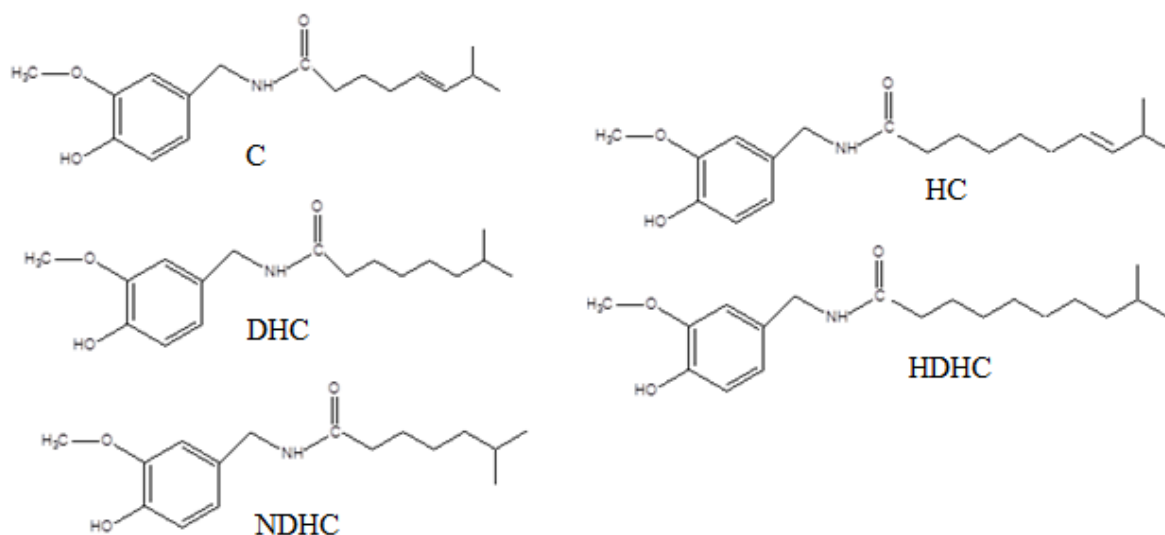


Figura IV.99. Capsaicina (C), dihidrocapsaicina (DHC), nordihidrocapsaicina (NDHC), homocapsaicina (HC) y homodihidrocapsaicina (HDHC)

Antes de llevar a cabo la determinación de C y DHC hay que realizar una extracción y limpieza del extracto. Se han descrito diversos métodos para la extracción de capsaicinoides en muestras de pimiento y pimentón, como la agitación magnética (MSE) combinada con el uso de SPE con cartuchos de C18 [53], fluidos supercríticos (SFE) y subcríticos [54], extracción asistida por microondas (MAE) [55] o extracción mediante ultrasonido (USE) [50]. También se han utilizado una gran variedad de disolventes como metanol [56], etanol [57], acetonitrilo [58], acetona [59,60], acetato de etilo [61], o mezclas de disolventes como acetona y éter etílico [62]. En este trabajo, se propone una extracción de C y DHC con acetonitrilo como extractante, y USE. Antes de la determinación mediante HPLC, es necesario aplicar un procedimiento de limpieza del extracto de la muestra mediante SPE con cartuchos de C18.

IV.4.1. Optimización de las condiciones cromatográficas

Teniendo en cuenta la bibliografía [63 – 65], se ha puesto a punto un método de cromatografía de fase inversa con detección por absorción UV. La selección de la longitud de onda se realizó de acuerdo a los datos referencia de la bibliografía [66] y teniendo en cuenta el espectro UV de la capsaicina y dihidrocapsaicina (figuras IV.100.- IV.101.). Se eligió 280 nm, que es la longitud de onda de máxima absorción de ambos compuestos.

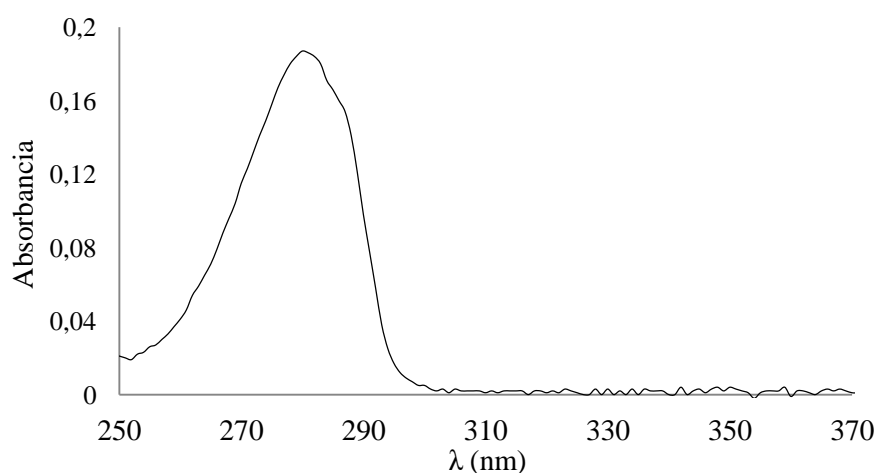


Figura IV.100. Espectro de absorción UV de capsaicina, 20 mg L⁻¹.

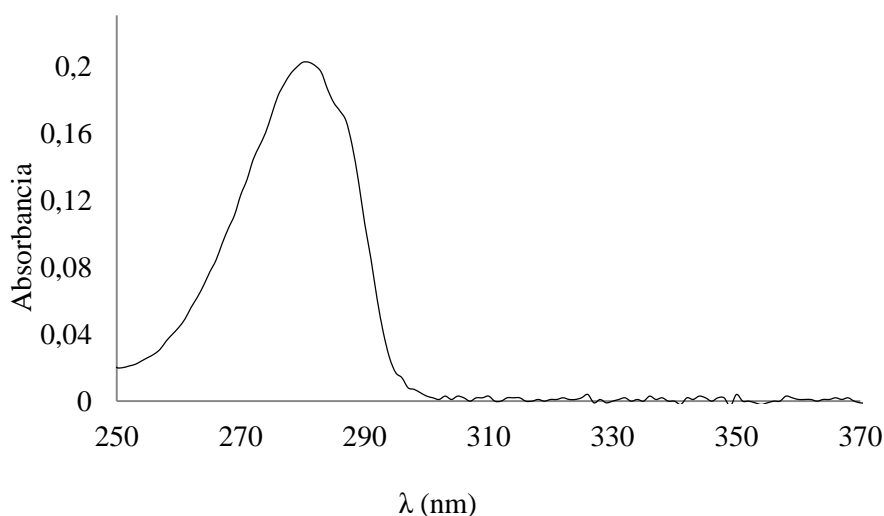


Figura IV.101. Espectro de absorción UV de dihidrocapsaicina, 20 mg L⁻¹.

De acuerdo con la bibliografía consultada, se ha utilizado una fase móvil acetonitrilo – agua, y una columna de C18. Para optimizar las condiciones cromatográficas se ha llevado a cabo un diseño experimental, en el los factores de respuesta utilizados han sido la resolución entre los picos, R , y la inversa del tiempo de retención del segundo pico, $1/t_{r2}'$. El diseño experimental ha constado de tres factores ($f=3$) codificados a dos niveles ($n=2$). Los factores son: el porcentaje de acetonitrilo en la fase móvil (x_1), el flujo de fase móvil (x_2) y la temperatura (x_3). Los niveles considerados para el primer factor fueron 50% y 40% de acetonitrilo, codificados como 1 y -1, respectivamente. Para el segundo factor se han tomado como niveles 1,2 y 1,0 mL min⁻¹, codificados como 1 y -1, respectivamente. Para el tercer factor los niveles codificados 1 y -1 fueron 50 y 40 °C, respectivamente. Para el nivel cero de cada factor se tomaron los valores medios de los dos niveles considerados, es decir 30% de acetonitrilo, 1,1 mL min⁻¹ y 45 °C. Se realizaron un total de 10 experimentos. Ocho de ellos definidos por el diseño, teniendo en cuenta los tres factores y los dos niveles de cada uno, $2^3 = 8$, más dos repeticiones del punto central con los valores del nivel cero. La tabla IV.36. muestra la matriz de diseño utilizada.

Tabla IV.36. Matriz de diseño para optimizar las condiciones cromatográficas

Experimento	Factor			Respuesta
	x_1	x_2	x_3	Y
A	1	1	1	Y_1
B	1	1	-1	Y_2
C	1	-1	-1	Y_3
D	1	-1	1	Y_4
E	-1	-1	1	Y_5
F	-1	1	-1	Y_6
G	-1	-1	-1	Y_7
H	-1	1	1	Y_8
I	0	0	0	Y_9
J	0	0	0	Y_{10}

x_1 = % ACN; x_2 = temperatura; x_3 =flujo fase móvil

Cada uno de los experimentos se realiza con muestras de disolución patrón, de concentración 25 mgL^{-1} , y se calculan los parámetros cromatográficos de interés, como son: el factor de retención, k' , y la resolución, R , obtenidos según las ecuaciones 4.67 y 4.68, respectivamente. El factor de retención proporciona información sobre el tiempo de elución. Cuando $k' < 1$, la elución es demasiado rápida y puede producirse el solapamiento de los picos de los analitos con los compuestos no retenidos. Si $k' > 20$, la retención es excesiva [67]. Por otra parte, la resolución es una medida cuantitativa del grado de separación de dos picos. El valor de R aceptado como referencia para una buena separación es 1,5 [68], aunque suele considerarse 2 como valor óptimo, para evitar que los picos se solapen debido a pequeños cambios en las condiciones cromatográficas.

$$k' = \frac{(t_r - t_0)}{t_0} \quad (4.67)$$

donde t_r es el tiempo de retención, t_0 es el tiempo muerto y

$$R = \frac{2[(t_r)_2 - (t_r)_1]}{W_1 + W_2} \quad (4.68)$$

donde W es la anchura de pico.

En los distintos cromatogramas registrados se obtienen valores de k' dentro del rango 3,98 – 7,56 para C y 6,09 – 11,92 para DHC. Estos valores de k' están, dentro del intervalo $1 < k' < 20$. Los valores de R varían entre 2,89 y 3,93.

Se buscan, las condiciones en las que R y $1/tr_2'$ presenten un valor máximo, lo que implica la mayor separación posible en menor tiempo de análisis. Los resultados obtenidos se muestran en la tabla IV.37. En la figura IV.102. se presentan los diagramas de superficie para las distintas respuestas elegidas, R, y $1/tr_2'$, en función de las variables estudiadas siendo A el porcentaje de acetonitrilo, B la temperatura y C el flujo de la fase móvil. En los diagramas de la izquierda se observa como R se hace mayor cuando aumenta el porcentaje de acetonitrilo (a) y en el caso (b) cuando disminuye. En el caso de la temperatura la variación no es demasiado importante en ninguno de los casos. En los diagramas centrales ocurre el mismo efecto contrapuesto con el porcentaje de acetonitrilo y la variación del flujo afecta en el sentido que a mayor flujo, mayor R y mayor $1/tr_2'$. Por último, en los diagramas de la derecha, se puede observar para la temperatura, en el caso (b) es más conveniente una temperatura más baja que en caso (a), mientras que el flujo es interesante mantenerlo alto en cualquier caso.

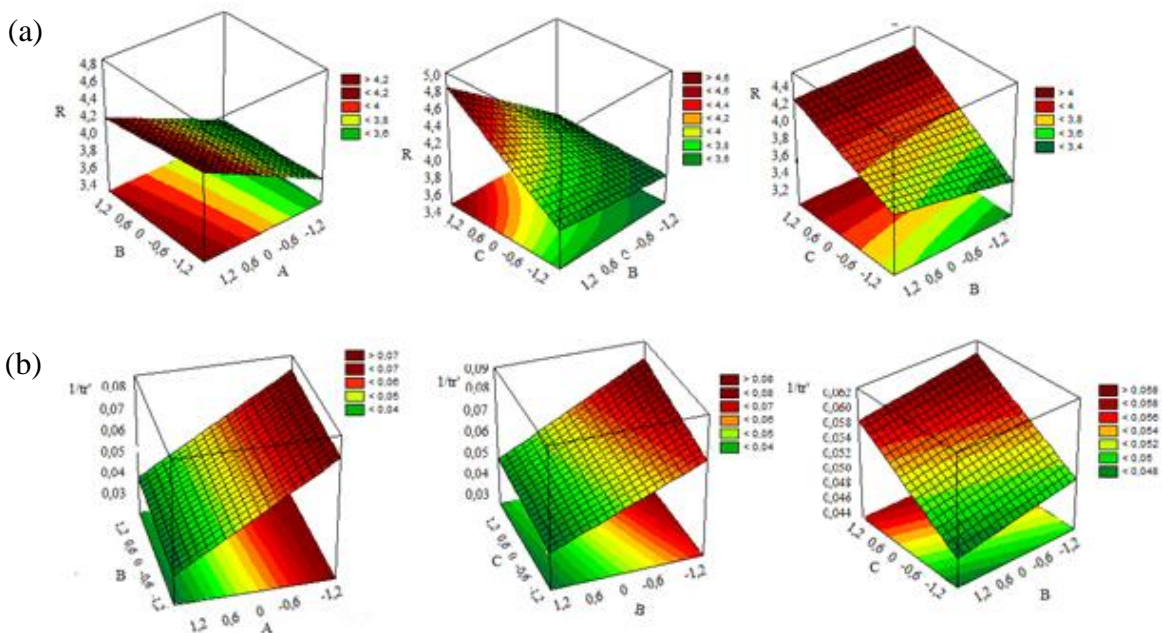


Figura IV.102. Superficie de respuesta para las distintas variables.
a) Resolución b) $1/tr_2'$. A: % acetonitrilo; B: temperatura; C: flujo.

Tabla IV.37. Resultados diseño experimental

Experimento	Respuesta	
	R	$1/t_{r2}'$
A	2,97	0,115
B	2,91	0,097
C	2,89	0,094
D	3,08	0,111
E	3,93	0,065
F	3,79	0,062
G	2,91	0,057
H	3,26	0,052
I	3,37	0,079
J	3,29	0,079

Es necesario llegar a una situación de compromiso que favorezca a ambas respuestas. Para comprobar si el efecto de los factores es significativo sobre las respuestas se han representado los diagramas de Pareto. En ambos casos, R (figura IV.103.a) y $1/t_{r2}'$ (figura IV.103.b) se observa que el porcentaje de acetonitrilo y el flujo producen una variación significativa. Cuando se considera la respuesta $1/t_{r2}'$, la temperatura juega un papel más importante que cuando se usa R. No obstante, en ambos casos, la influencia de la temperatura no es significativa así como la combinación de ésta con los demás factores.

En la figura IV.104. se representan la variación de los valores promedio de R y $1/t_{r2}'$, junto con su desviación estándar para los tres factores. En el caso del porcentaje del acetonitrilo se produce una disminución de la respuesta R y un aumento para el caso de $1/t_{r2}'$. La combinación de ambas respuestas presenta un máximo para el valor central del diseño. La temperatura tiene una ligera influencia negativa en el caso de R, y positiva para $1/t_{r2}'$. La combinación de las dos respuestas, prácticamente, no sufre variación para los niveles considerados en el diseño. El flujo de fase móvil da lugar a un aumento en las dos respuestas, y también en la combinación de respuestas.

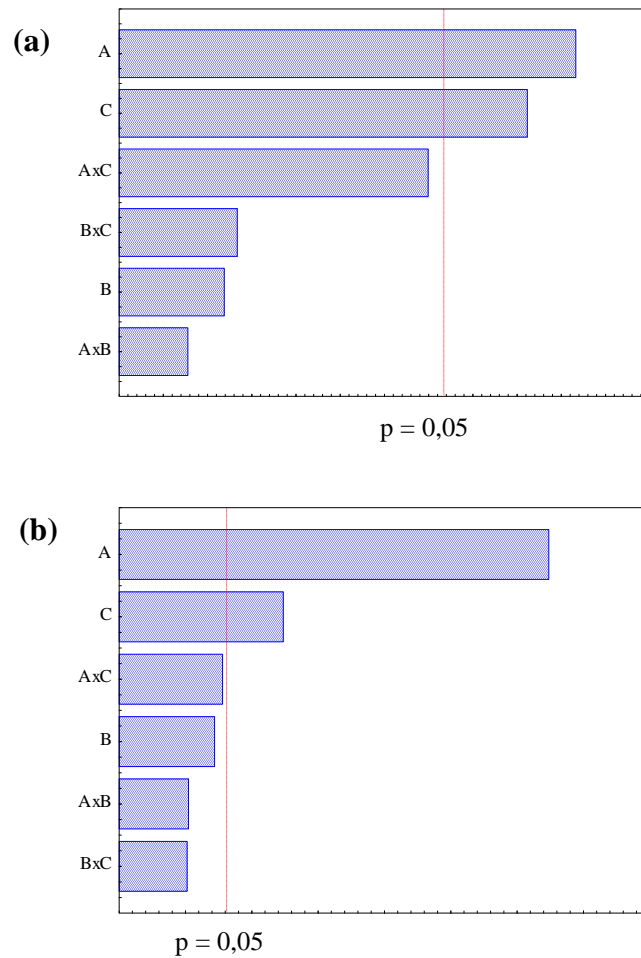


Figura IV.103. Diagrama de Pareto. a) R; b) $1/tr_2'$.

A: % acetonitrilo; B: temperatura; C: flujo.

Cuando se considera la respuesta R, los valores mayores se obtienen cuando disminuye el porcentaje de acetonitrilo, al contrario que $1/tr_2'$, que disminuye en esos casos. Si analizamos el efecto de la temperatura no se observa una variación tan drástica como con el porcentaje de acetonitrilo, pero la tendencia es similar, cuando disminuye la temperatura encontramos valores de R mayores y ligeramente menores de $1/tr_2'$. En el caso del flujo, cuando trabajamos a flujos bajos, R toma los valores más bajos al igual que $1/tr_2'$, aunque no existen tanta oscilación en este último factor..

Se combinan ambas respuestas en una función deseada para obtener los valores óptimos de los factores estudiados para las dos variables al mismo tiempo. Se llega a una situación de compromiso en la que las condiciones óptimas del porcentaje de acetonitrilo y la temperatura son, 43% acetonitrilo y 43°C. En el caso del flujo, el

máximo para ambas respuestas se obtiene para un flujo de $1,2 \text{ mL min}^{-1}$, y es previsible un aumento con el mismo. Un aumento de flujo implica un aumento de la presión de trabajo en el sistema cromatográfico, y por cuestiones de tipo práctico se elige un flujo de $1,1 \text{ mL min}^{-1}$.

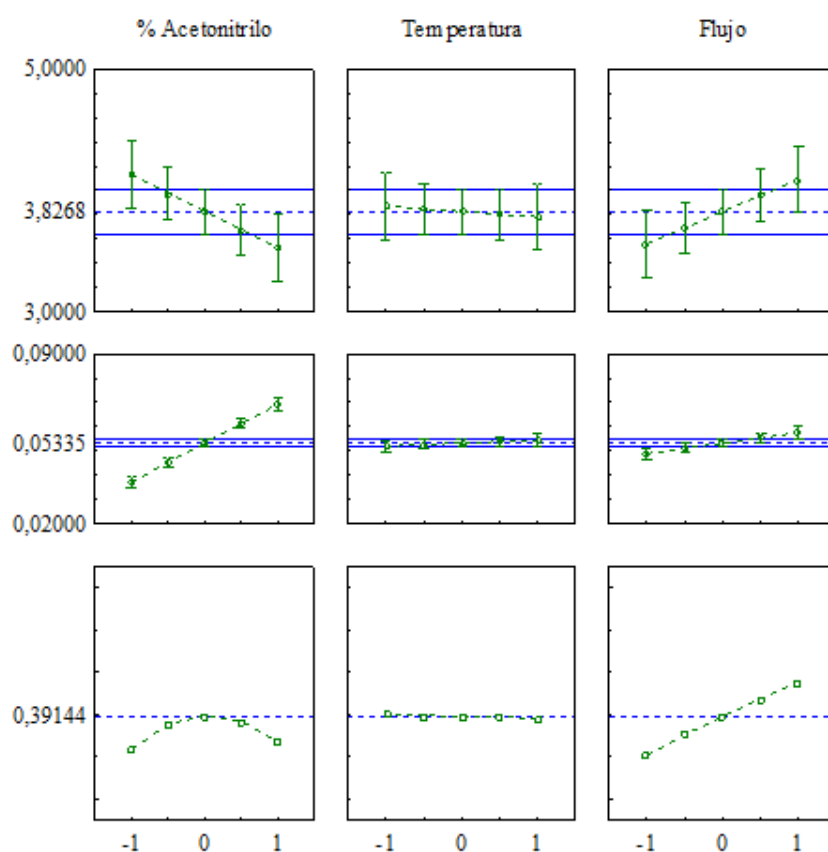


Figura IV. 104. Variación de las respuestas individuales y función deseada en función de los factores considerados en el diseño experimental.

En la figura IV.105. se presentan las distintas funciones de deseo para las variables estudiadas. Cuando se estudia la función de deseo en el caso del % de acetonitrilo y la temperatura, se observa como la situación más favorable es cuando se tienen % de acetonitrilo y temperatura media. Cuando se representa la función de deseo para las variables flujo y % de acetonitrilo, los valores más altos de dicha función se dan cuando se trabaja a flujos altos y % medio de acetonitrilo, mientras que si se estudia el flujo y la temperatura, los valores más altos de la función de deseo se observan a flujos altos independientemente de la temperatura, que no afecta tanto como el flujo.

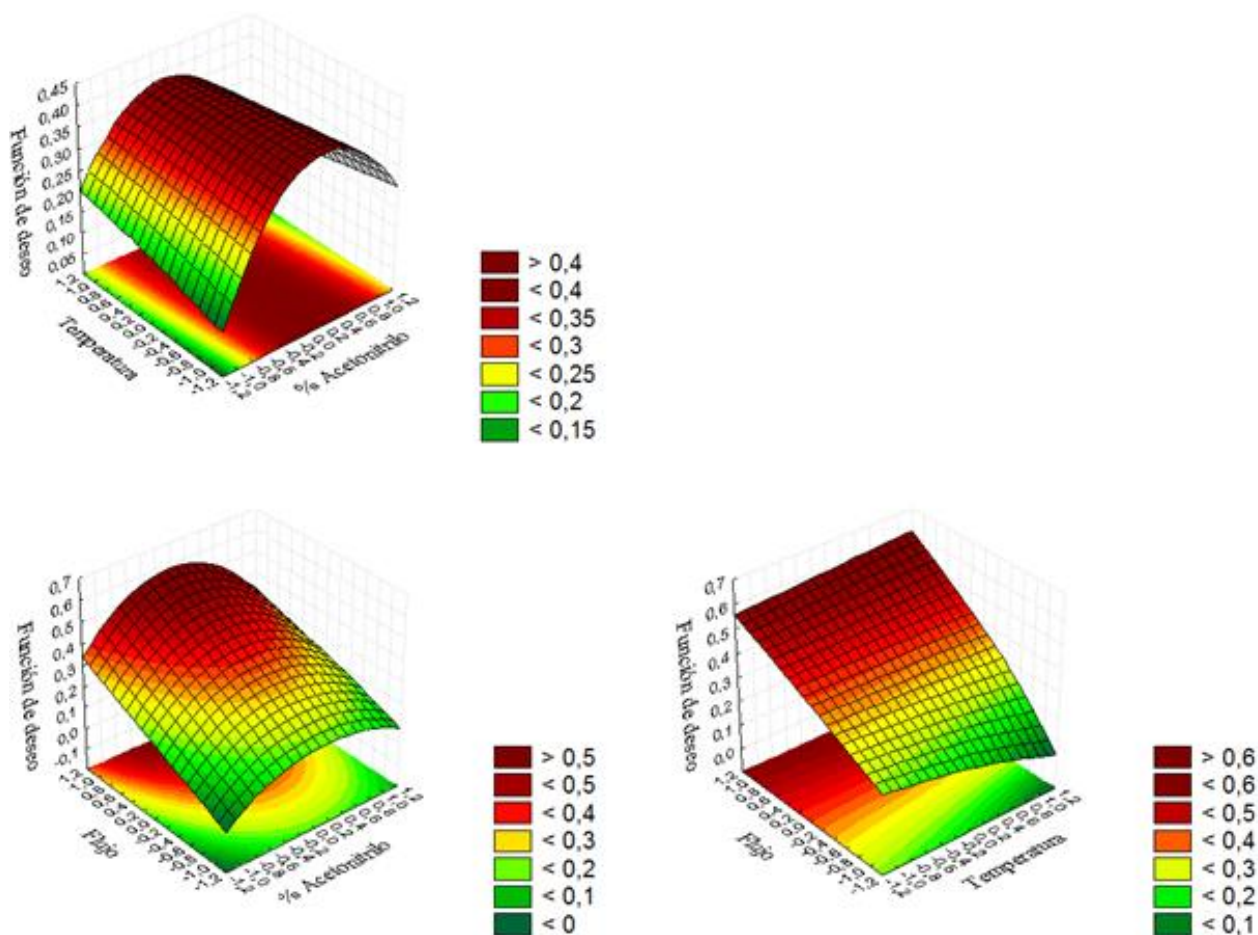
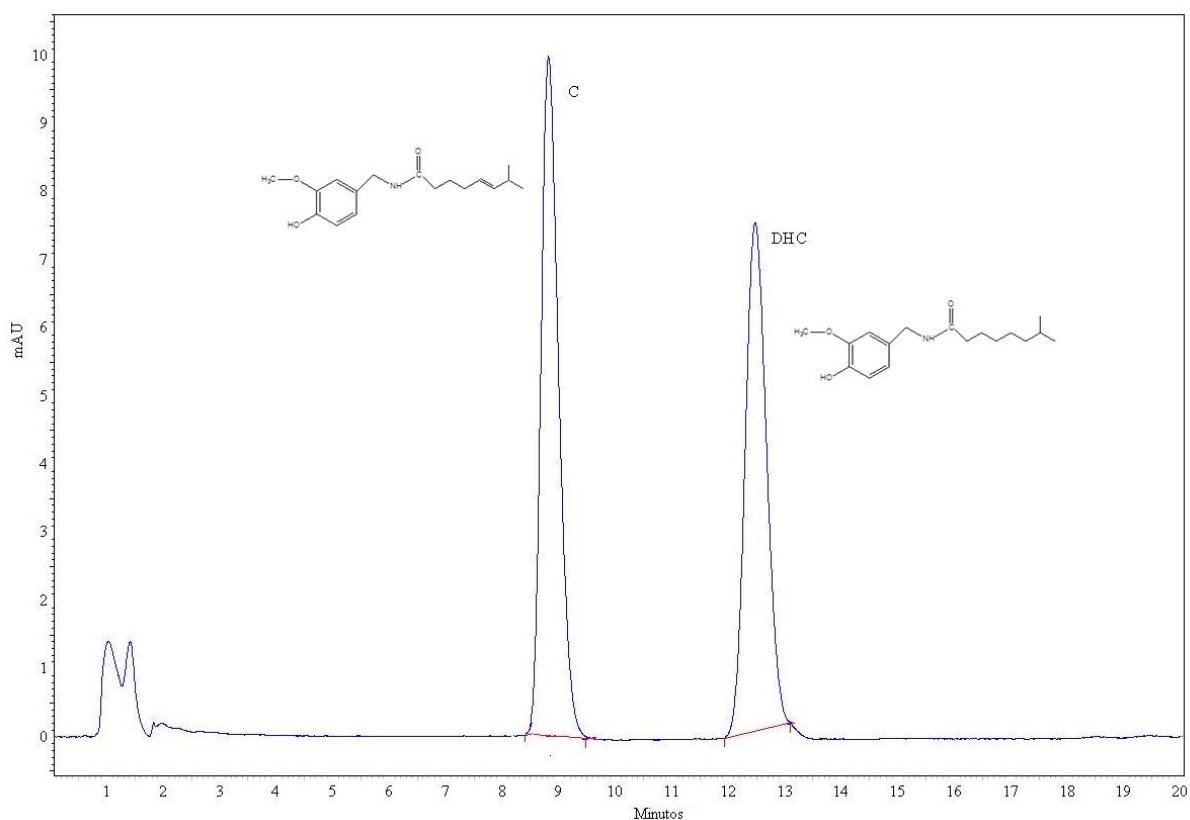


Figura IV.105. Diagramas de superficie de la función deseada

Las condiciones cromatográficas óptimas elegidas a partir del diseño experimental se recogen en la tabla IV.38. En la figura IV.106. se muestra un cromatograma de una disolución conteniendo concentraciones de 25 mg L⁻¹ de C y DHC. Se comprueba que los picos de los compuestos están completamente separados. Los tiempos de retención son 8,7 y 12,6 min para la C y DHC, respectivamente. Con estas condiciones cromatográficas se consigue una separación adecuada de los dos analitos considerados, y además la retención es suficiente para que no haya solapamiento con los picos de compuestos no retenidos que pueden aparecer al principio del cromatograma cuando se analicen las muestras de pimentón.

Tabla IV.38. Condiciones cromatográficas para la determinación de C y DHC en muestras de pimentón

Fase móvil	Acetonitrilo/agua (63:37)
Fase estacionaria	C18
Flujo de la fase móvil	1,1 mL min ⁻¹
Volumen de inyección	20 µL
Temperatura de la columna	43°C
Longitud de onda de detección	280 nm

**Figura IV.106.** Cromatograma disolución patrón de 25 mg L⁻¹ de capsaicina (C) y dihidrocapsaicina (DHC).

IV.4.1.1. Asimetría y factor de cola

Los picos cromatográficos deben tener forma gaussiana, pero en muchos casos se producen desviaciones del comportamiento ideal. La asimetría (As) y el factor de

cola del pico (PTF) son los parámetros que se utilizan para evaluar la simetría de un pico cromatográfico.

La asimetría del pico (A_s) se calcula mediante la ecuación 4.69. Los segmentos A y B se miden a una altura igual al 10% de la altura total pico.

$$A_s = \frac{B}{A} \quad (4.69)$$

B es la distancia desde la bisectriz del pico hasta la cola del mismo y A la distancia entre el frente del pico y dicha bisectriz. Se considera que la forma de un pico es simétrica si A_s toma valores comprendidos entre 1,0 y 1,05, siendo 1,2 el valor máximo aceptable [68].

Para calcular PTF se utiliza la expresión:

$$PTF = \frac{B + A}{2A} \quad (4.70)$$

En este caso los valores de A y B se miden a una altura igual al 5% de la altura total de pico. Este factor también debe tener un valor próximo a 1 para considerar que la simetría de pico es adecuada.

Los valores de A_s y PTF se calcularon para los picos obtenidos para una muestra de pimentón con las condiciones cromatográficas óptimas. En el caso de la C se obtiene un valor de 1,1 para A_s y 0,97 para PTF. Cuando se trata de la DHC los valores son de 0,96 y 1,02 para A_s y PTF, respectivamente. Los valores obtenidos cumplen los requisitos de calidad de un pico cromatográfico.

IV.4.2. Optimización de las condiciones de extracción

Como ya se indicó en el apartado IV.4, teniendo en cuenta la bibliografía consultada se seleccionó el acetonitrilo como disolvente, para utilizarlo en la extracción con ultrasonido. Se han estudiado las condiciones de extracción de C y DHC en pimentón. Las variables que se han considerado son: volumen de extractante, tiempo de extracción y temperatura de extracción. La optimización se ha realizado variable a

variable de manera independiente y los resultados obtenidos se presentan en las figuras IV.104 – IV.106. La técnica de extracción elegida en este caso no ha sido una variable en el estudio si no que se ha elegido en base a la bibliografía existente y a las ventajas que ofrece con respecto a otros métodos. La extracción USE se basa en la energía ultrasónica para conseguir mayor eficiencia en el contacto entre la muestra y el disolvente [69]. Este método que permite la extracción de matrices sólidas con disolventes líquidos, sin tener que emplear un volumen excesivo de disolvente, y utilizando un tiempo de extracción menor que en el caso de la extracción Soxhlet [70] o métodos clásicos como la maceración [71]. Pastor et al. [72] han obtenido mediante USE recuperaciones similares a las logradas con la MAE. Estas ventajas sumadas a que numerosos autores [50, 54], en trabajos previos, han recurrido a USE para la extracción de capsaicinoides han servido para decidir sobre el método de extracción. La elección de acetonitrilo como disolvente para la extracción, está basada en varios trabajos donde se concluye que el uso de acetonitrilo disminuye la cantidad de interferencias provocadas por los carotenoides, presentes en la muestra en cantidades importantes [73,74]. Aunque algunos autores proponen que la adición de pequeñas porciones de agua al volumen total de disolvente favorece la extracción [75], no se ha realizado, ya que el capsaicinoide menos polar, o sea, la dihidrocapsaicina no se extraía correctamente. Este hecho también ha sido confirmado por otros autores [76].

En primer lugar se ha considerado el volumen de extractante. Por lo general, en las extracciones sólido-líquido una disminución de la relación masa/volumen, favorece la extracción, ya que hay un mismo volumen para extraer una cantidad menor de soluto [77]. En este caso se consideró la extracción de 1 g de muestra con 5, 10, 15 y 20 mL de acetonitrilo.

La extracción se realizó con ultrasonido a 50°C, durante un tiempo de 20 minutos. Estas condiciones que se mantuvieron fijas en los diversos ensayos realizados con volúmenes diferentes de disolvente extractante. Los resultados obtenidos se muestran en la figura IV.107. Como se puede apreciar, la extracción máxima de capsaicina se obtiene a partir de un volumen de 10 mL de extractante, con una pequeña disminución para volúmenes superiores. En el caso de la dihidrocapsaicina se obtienen

rendimientos para un volumen de 20 mL de acetonitrilo. En función de estos resultados, los experimentos posteriores se hicieron utilizando este volumen de extractante.

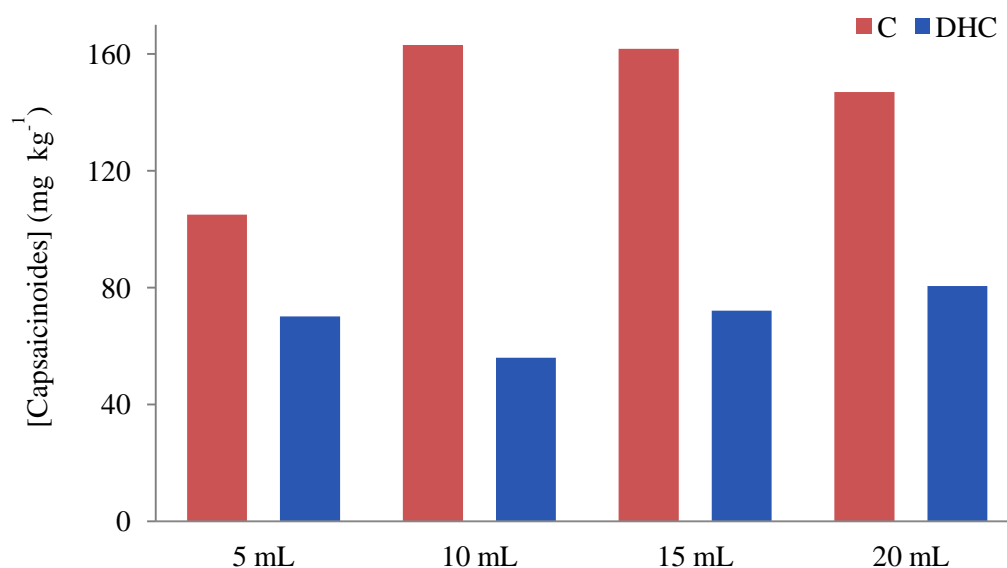


Figura IV.107. Concentración de capsaicinoides extraídos (mg kg⁻¹) en función del volumen de extractante

Para optimizar el tiempo de extracción se consideraron 5, 15, 20 y 30 min, manteniendo constante un volumen de 20 mL de acetonitrilo, para un peso de la muestra 1 g, y una temperatura de 50°C. El tiempo es una variable relevante a estudiar, ya que tiempos cortos pueden dar lugar a que la extracción no sea completa, y en algunos casos, si el tiempo es demasiado largo se puede producir algún tipo de degradación en la muestra. Los resultados obtenidos se muestran en la figura IV.108. Como se puede observar, a partir de 15 min la extracción de capsaicina alcanza el valor máximo. Para la dihidrocapsaicina, el valor máximo se alcanza a partir de los 20 min de extracción. Se eligió un tiempo de 20 min para asegurar la extracción de los dos compuestos. Los experimentos posteriores para optimizar la temperatura se llevaron a cabo utilizando un volumen de 20 mL de disolvente y un tiempo de extracción de 20 min.

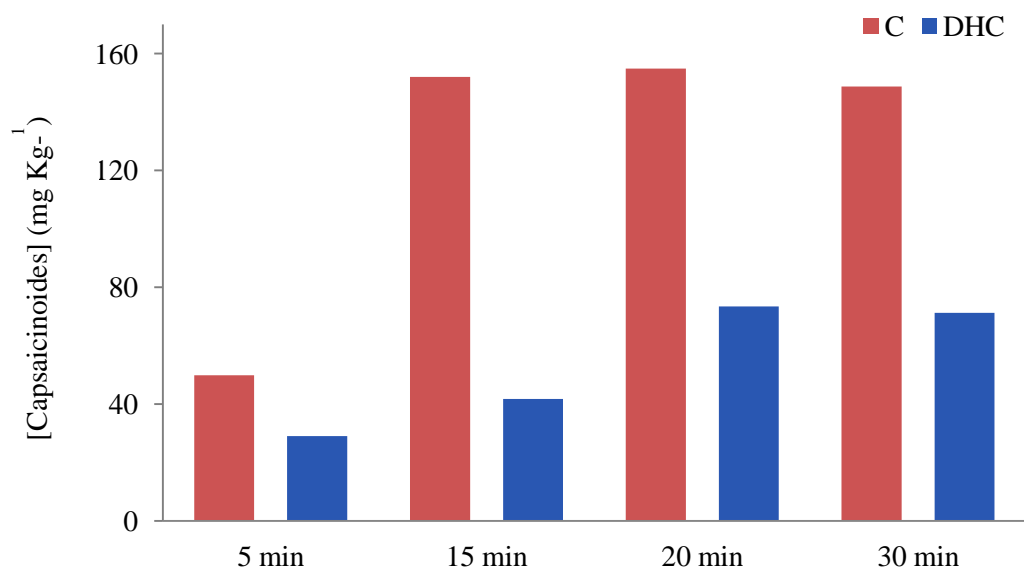


Figura IV.108. Concentración de capsaicinoides extraídos (mg kg^{-1}) en función del tiempo de extracción.

Una vez fijados el volumen de extractante y el tiempo de extracción, se optimiza la temperatura. Se consideraron las temperaturas 30, 40, 50 y 60°C, manteniendo fijo un volumen de acetonitrilo y el tiempo. Los resultados obtenidos se presentan en la figura IV.109. Como puede apreciarse la cantidad máxima, tanto de capsaicina como de dihidrocapsaicina se obtiene con una temperatura de 50 °C. A temperaturas superiores se aprecia una disminución de la cantidad extraída de ambos compuestos. Este efecto puede deberse a una descomposición térmica de los de C y DHC, que da lugar a una bajada del rendimiento de extracción. Aunque algunos autores [78] han descrito que los capsaicinoides pueden soportar temperaturas más elevadas sin que se produzca una degradación de los mismos, se ha considerado que la temperatura más adecuada para llevar a cabo la extracción es de 50°C. De esta forma se previene la posible degradación térmica de los capsaicinoides.

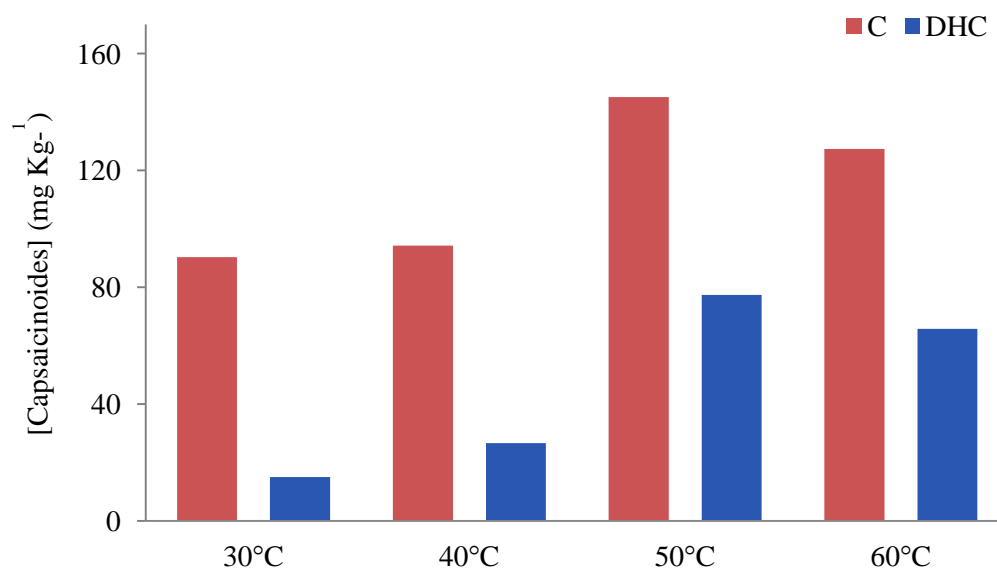


Figura IV.109. Concentración de capsaicinoides extraídos (mg kg^{-1}) en función de la temperatura de extracción.

Una vez realizada la extracción es necesario llevar a cabo la limpieza del extracto. Para ello se han utilizado SPE. Considerando los datos encontrados en la bibliografía [79] se utilizaron cartuchos de C18 y acetonitrilo como disolvente para eluir los capsaicinoides retenidos en el cartucho. Teniendo en cuenta el trabajo realizado por Attuquayefio et al. [79], sobre la extracción de capsaicinoides en muestras de pimentón y de pimiento fresco, se utilizaron cartuchos SPE, con 200 mg de C18. En ese trabajo la muestra se eluye con 4 mL de acetonitrilo y 1 mL de acetonitrilo al 1% en ácido acético para evitar la turbidez. En nuestro caso, no se empleó acético puesto que no se apreciaba turbidez, pero sí se microfiltró la muestra con un filtro de nylon de $0,45 \mu\text{m}$, llevándola a un volumen final de 5 mL, antes de la inyección en el cromatógrafo.

Tras decidir las condiciones de extracción, realizar la limpieza y microfiltrar, se comprueba si existe efecto matriz, y se valida el método en términos de veracidad y precisión.

IV.4.3. Efecto matriz

Para estudiar la presencia de efecto matriz, se prepararon rectas de calibrado externo y de adición patrón sobre una muestra de control, representando el área de los picos cromatográficos frente a la concentración de los patrones o las adiciones. En ausencia de efecto matriz, las pendientes deben ser estadísticamente iguales. Los resultados se calculan según las ecuaciones 4.10-4.12. Los grados de libertad se calculan de acuerdo a la ecuación Welch-Satterthwaite (ecuación 4.13). Cuando el valor de t_{cal} es menor que el de t_{tab} (v , 95%) se considera que no existe efecto matriz. En la tabla IV.39. se pueden observar estos resultados.

Tabla IV.39. Efecto matriz en la determinación de capsaicinoides.

Analito	b_{AP}	s_{bAP}	b_{CE}	s_{bCE}	t_{cal}	t_{tab} (95%)
Capsaicina	36501,9	1913,0	31402,6	553,32	2,94	4,30
Dihidrocapsaicina	26904,3	2012,1	30071,0	108,78	1,67	1,96

b : pendiente de la recta de calibrado. s_b : desviación estándar de la pendiente. a : ordenada en el origen de la recta de calibrado. s_a : desviación estándar de la ordenada en el origen. t_{cal} : t calculada. t_{tab} : t tabulada (AP: adición patrón; CE: calibrado externo).

Los valores de t_{cal} son menores que los de t_{tab} en ambos casos. De esta forma se comprueba que los valores de las pendientes de las rectas de calibrado externo y de adición patrón de la capsaicina y dihidrocapsaicina son estadísticamente iguales. Por lo tanto, se puede afirmar que no existe efecto matriz. En las figuras IV.110. y IV.111. se muestran las rectas de calibrado y adición patrón para ambos compuestos. Teniendo en cuenta la ausencia de efecto matriz, las determinaciones de capsaicina y dihidrocapsaicina por HPLC se realizaron mediante calibrado externo.

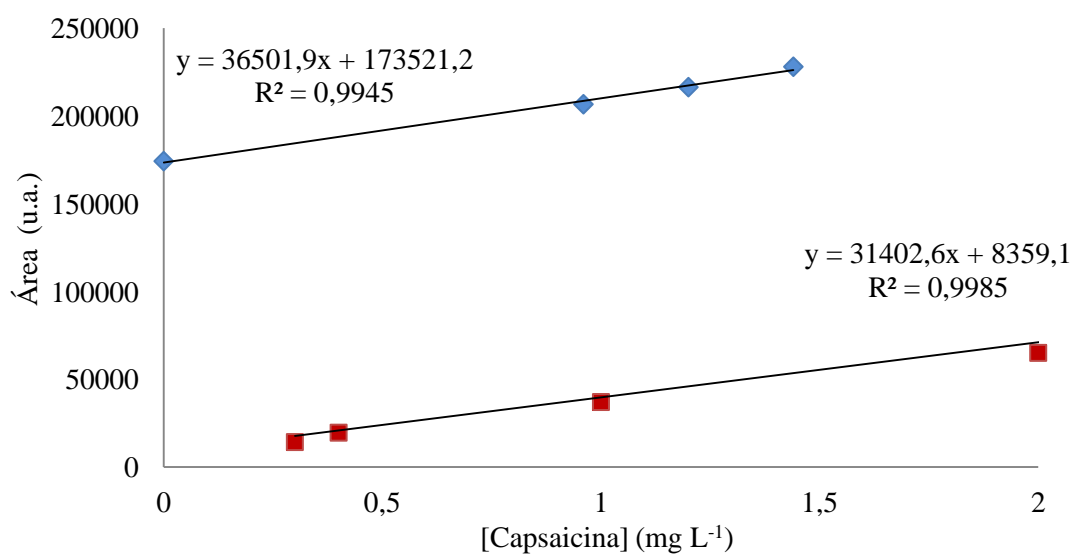


Figura IV.110. Recta de calibrado externo (CE ■) y adición patrón (AP◆) de capsaicina.

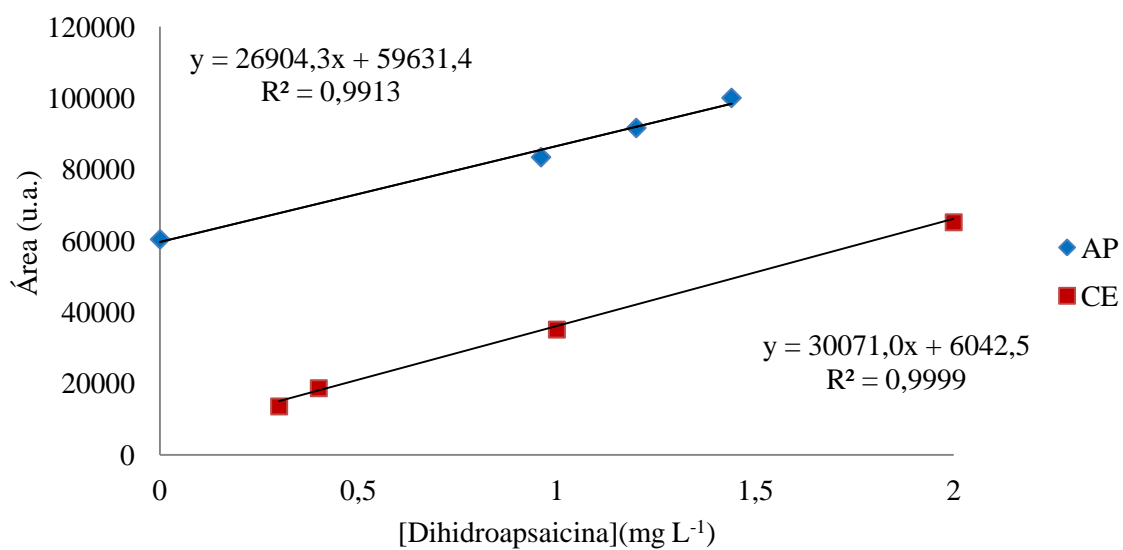


Figura IV.111. Recta de calibrado externo (CE ■) y adición patrón (AP◆) de dihidrocapsaicina.

IV.4.4. Parámetros de calidad del método

IV.4.4.1. Veracidad

La veracidad se comprobó mediante ensayos de recuperación. Se analizaron muestras fortificadas en tres niveles de adición (tabla IV.40.), realizando tres repeticiones en cada nivel. Se calculó la incertidumbre expandida, y utilizó un factor de cobertura igual a 2 [5].

Tabla IV.40. Niveles de adición recuperación capsaicinoides.

Nivel de adición	1	2	3
Capsaicina (mg L ⁻¹)	3,2	4	4,8
Dihidrocapsaicina (mg L ⁻¹)	3,2	4	4,8

Los valores de recuperación medios así como la incertidumbre expandida, $U(\bar{R}_i)$, para la C y DHC se recogen en la tabla IV.41. En el caso de C la recuperación es del 96% y el valor de $U(\bar{R}_i)$ 15. Para la DHC los valores de recuperación y $U(\bar{R}_i)$ son 97% y 14, respectivamente. La desigualdad $|\bar{R}_i - 100| \leq U(\bar{R}_i)$ se cumple en todos los casos, y los valores de recuperación obtenidos con sus incertidumbres expandidas, contienen al 100%, pudiendo afirmar que es un método veraz.

Tabla IV.41. Recuperaciones de C y DHC.

Analito	Recuperación (%)
	(R ± U)
Capsaicina	96 ± 15
Dihidrocapsaicina	97 ± 14

U: incertidumbre expandida.

IV.4.4.2. Precisión

La precisión del método se ha estudiado considerando la repetitividad y la precisión intermedia. Se analizó una muestra de control en condiciones de repetitividad y precisión intermedia [10]. Mediante un ANOVA se ha calculado la precisión intermedia y la repetitividad que se muestran en la tabla IV.42. en términos de desviación estándar relativa. También se han incluido los valores Horwitz de referencia para las concentraciones del orden correspondiente. Los valores de Horwitz y de la AOAC para concentraciones de 10 ppm son 11,3 y 7,3, respectivamente y 8 y 5,3 para concentraciones del orden de 100 ppm. En todos los casos, los resultados obtenidos están dentro de los rangos dados por Horwitz y por la AOAC [11, 81]. Se incluye, además, el parámetro Horrat calculado según las ecuaciones 4.18 y 4.19. Como los valores son inferiores a 1, se puede considerar que la precisión es aceptable.

Tabla IV.42. Repetitividad y precisión intermedia para la determinación de C y DHC en pimentón mediante HPLC. Criterio de aceptación de Horwitz, AOAC, Horrat

Analito	Repetitividad (%)	Precisión Intermedia (%)	Horwitz %	AOAC %	Horrat
C	3,81	4,44	16	11	0,39
DHC	3,59	4,27	16	11	0,38

IV.4.4.3. Calibración, linealidad y límites de detección y cuantificación

Se realizó un calibrado externo para la determinación de capsaicinoides en muestras de pimentones con patrones desde 0,3 mg L⁻¹ a 20 mg L⁻¹. Se utiliza el área de pico cromatográfico como señal analítica, que se representa frente a la concentración para obtener las rectas de calibrado. La tabla IV.43 incluye los datos correspondientes a las pendientes, las ordenadas en el origen, y sus respectivas desviaciones estándar; también se han incluido los valores de los coeficientes de determinación. Las figuras IV.112 – IV.113. muestran las rectas de calibrado correspondientes a la capsaicina y dihidrocapsaicina, respectivamente.

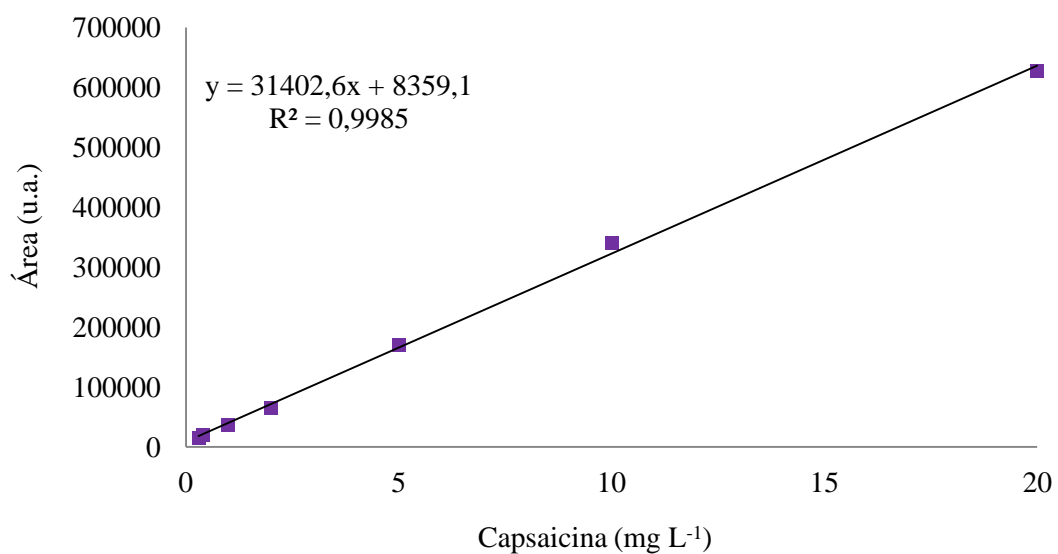


Figura IV.112. Recta de calibrado externo para la determinación de capsaicina.

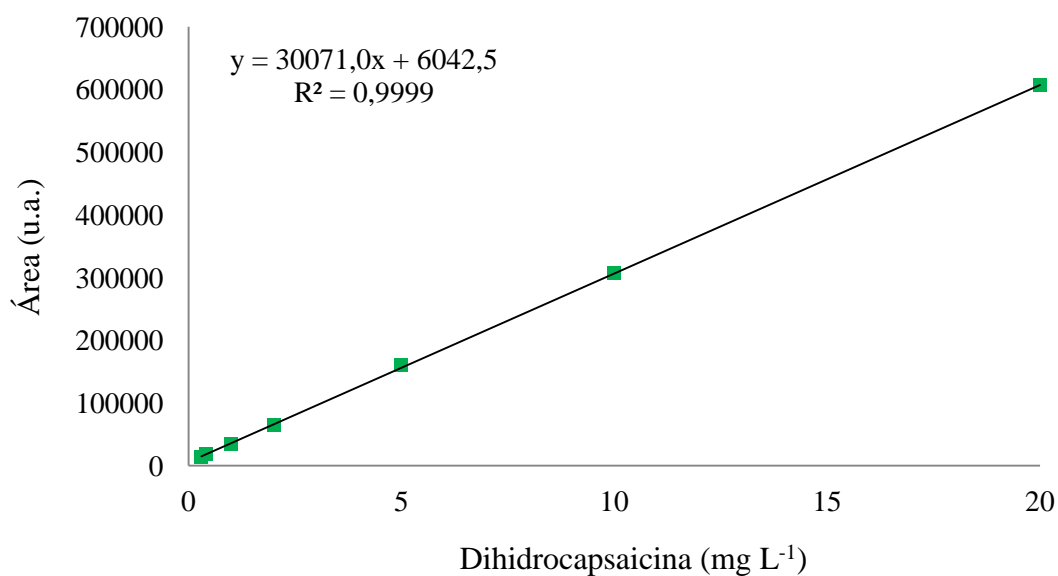


Figura IV.113. Recta de calibrado externo para la determinación de dihidrocapsaicina.

Tabla IV.43. Parámetros de calibración para la determinación de C y DHC en muestras de pimentón picante por HPLC.

Analito	Rango concentración (mgL ⁻¹)	b	s _b	a	s _a	S _{y/x}	R ²
C	0,3-20	31402,6	553,319	8359,1	4815,8	9840,6	0,9985
DHC	0,3-20	30071,0	108,784	6042,5	946,8	1934,7	0,9999

b: pendiente de la recta de calibrado. s_b: desviación estándar de la pendiente. a: ordenada en el origen de la recta de calibrado. s_a: desviación estándar de la ordenada en el origen. S_{y/x}: error típico de la recta de calibrado. R²= coeficiente de determinación.

Otros parámetros de calidad son la linealidad y los límites de cuantificación (LOQ) y detección (LOD). Los valores de LOD y LOQ se calculan a partir de la relación señal ruido [68], considerando una relación señal/ruido 3:1 y 10:1 para los LOD y LOQ, respectivamente. En la tabla IV.44. se recogen los valores calculados para la C y DHC.

Tabla IV.44. LOD y LOQ, mg L⁻¹, para C y DHC

Analito	LOD	LOQ
C	0,06	0,2
DHC	0,12	0,4

La linealidad se establece a partir de la pendiente de la recta de calibrado y su desviación estándar, según la ecuación. 4.22 [12]. En el caso de la capsaicina la linealidad es de un 98,4%, y en el caso de la dihidrocapsaicina es igual a 99,4%. También se puede estudiar la linealidad gráficamente, representando el factor de respuesta frente a la concentración. Se define como rango lineal aquél en el que los valores de señal/concentración se mantienen entre un 95-105% del valor de la mediana de estos valores [82]. Como se observa en la figura IV.114., para la capsaicina, el rango lineal varía en el intervalo 0,3 – 20 mg L⁻¹, ya que todos los puntos están dentro de los límites de confianza. En el caso de la dihidrocapsaicina (figura IV.115.) el rango es lineal entre 0,3 y 20 mg L⁻¹.

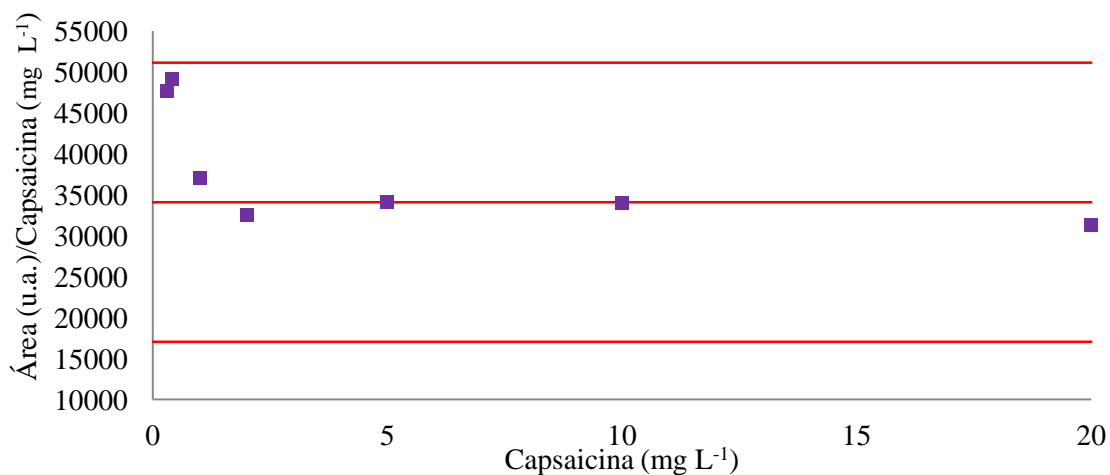


Figura IV.114. Rango lineal capsaicina.

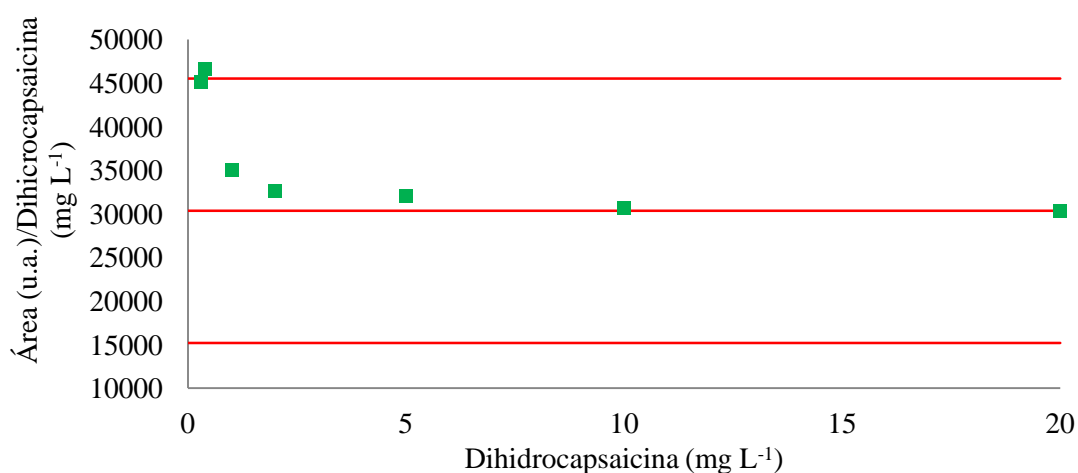


Figura IV.115. Rango lineal dihidrocapsaicina.

IV.4.5. Determinación de capsaicinoides en pimentones picantes.

El método puesto a punto se ha aplicado al análisis de muestras de pimentón picante. En la tabla A1.3. del anexo 1 se muestran los resultados obtenidos. Los valores que se encuentran bajo el LOQ no han sido eliminados de la tabla de acuerdo a las recomendaciones de la IUPAC [83]. En la tabla IV.45. se muestran los valores de la mediana, así como los máximos y mínimos obtenidos. Además de la concentración de C y DHC se ha calculado el grado de picor de la muestra en unidades Scoville [84].

La relación que existe entre la concentración de capsaicina y las unidades Scoville (SHU) es la siguiente [85]:

$$SHU = [Capsaicina] \left(\frac{g}{g}\right) \times 16,1 \times 10^6 \quad (4.71)$$

Los valores de la mediana para las muestras del origen MU para las concentraciones de C y DHC son mayores que los valores de los pimentones procedentes de Extremadura. Los valores de las medianas de la concentración de DHC son más parecidos entre orígenes que los valores de C. En el caso de los valores mínimos, las muestras procedentes de Extremadura presentan valores más bajos, tanto en el caso de C como de DHC. El valor mínimo de C para Extremadura es 26,4 mg kg⁻¹ y el de DHC 17,2 mg kg⁻¹. Los pimentones procedentes de Murcia son los que presentan valores máximos más elevados para los dos capsaicinoides, 919,8 y 476,4 mg kg⁻¹ para C y DHC, respectivamente. La dispersión en los valores de C de Murcia es mayor que en el caso de Extremadura.

Tabla IV.45. Valores de mediana, mínimo y máximo (mg kg⁻¹) de C y DHC y SHU para pimentón picante

Variable	MURCIA	EXTREMADURA
	Picante	
C	158,6	104,2
	(90,8-919,8)	(26,4-278,1)
DHC	114,1	74,3
	(27,5-476,4)	(17,2-256,1)
SHU	2254,0	1677,5
	(1463-14810)	(424-4477)

Valores mínimos y máximos entre paréntesis. C: capsaicina, DHC: dihidrocapsaicina; SHU: unidades Scoville.

El parámetro SHU se calcula en función de la concentración de capsaicina (ecuación 4.71), por lo que la mayor mediana y el valor máximo se corresponden con los pimentones del origen Murcia. El valor mínimo de SHU pertenece a una muestra

procedente de Extremadura, al igual que lo hacía el valor mínimo de C. En el caso de Murcia, encontramos un intervalo de valores muy amplia que va desde 1463 hasta 14810 unidades SHU. En el caso de los pimentones picantes de Extremadura el rango de SHU comprende valores desde 424 hasta 4477. A pesar de las diferencias entre los orígenes, en casi todos los casos nos encontramos con unidades SHU correspondientes al sabor picante, como se puede ver en la tabla IV.46.

La figura IV.116.muestra un cromatograma de una muestra de pimentón picante donde se observan los picos de C y DHC a los tiempos de retención esperados para las condiciones cromatográficas dadas. En la parte inicial del cromatograma aparece un conjunto de picos que corresponden a compuestos no retenidos. Se aprecia una buena resolución entre los picos de los analitos, capsaicina y dihidrocapsaicina, y los otros picos que aparecen en el cromatograma.

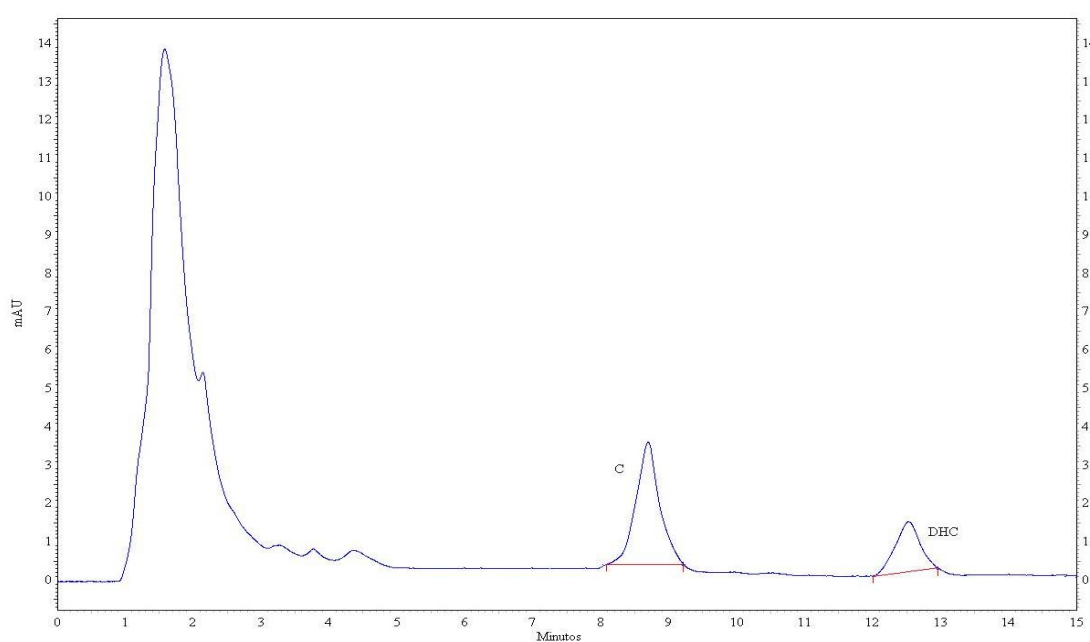


Figura IV.116. Cromatograma de una muestra de pimentón picante.

Condiciones cromatográficas tabla IV.3.

Los resultados obtenidos coinciden con los que aparecen en la bibliografía para muestras similares. Barbero et al. [55] determinaron capsaicinoides en pimientos mediante extracción por microondas y HPLC con detector de fluorescencia encontrando valores en torno a los 200 mg kg⁻¹ de capsaicina y 80 mg kg⁻¹ de dihidrocapsaicina. Perucka et al. [86] determinan mediante HPLC, además de carotenoides, los capsaicinoides de muestras de pimientos del género *Capsicum annuum* L., obteniendo valores medios totales capsaicina más dihidrocapsaicina de 700 mg kg⁻¹. Juangsamoot et al. [53] determinaron capsaicina y dihidrocapsaicina por RP-HPLC, llevando a cabo la extracción mediante agitación magnética y limpieza del extracto con cartuchos C18. En este caso se analizaron pimientos chiles picantes y se obtuvieron concentraciones de capsaicinoides dentro del rango de 1758-7068 mg kg⁻¹, lo que se corresponde con valores de en torno a 25.000-80.000 SHU. En cuanto al pimentón, Daood et al. [54] estudiaron la extracción de capsaicinoides en pimentones húngaros y encontraron valores totales en torno a 500 mg kg⁻¹, que implican pimentones con 8000 SHU; este dato coincide con los valores mostrados en la tabla IV.46. López, Simal y Vázquez [87] determinaron la concentración de capsaicina en pimentones españoles mediante espectroscopía derivada, obteniendo valores que oscilan entre 80 y 400 mg kg⁻¹, teniendo en cuenta que la recuperación del método fue del 84%. En este estudio de pimentones españoles, se obtienen concentraciones similares. En el caso de las unidades SHU el rango es amplio y se encuentran valores correspondientes a pimientos muy ligeramente picante, chile rocotillo y pimiento de Padrón, en el caso de los más picantes. Hay que recordar que la escala Scoville está basada en parámetros organolépticos y está construida en base a la capsaicina pura [85].

Tabla IV.46. Escala Scoville.

Unidades Scoville	Producto
15.000.000– 16.000.000	Capsaicina pura
8.600.000–9.100.000	Varios capsaicinoides: homocapsaicina, homodihydrocapsaicina y nordihydrocapsaicina
2.000.000–5.300.000	Nivel estándar del aerosol de pimienta en EE.UU., munición irritante
2.200.000	Carolina reaper
1.300.000–2.000.000	Naga viper, Trinidad Scorpion Butch T
855.000–1.041.427	Naga Jolokia
350.000–580.000	Habanero savinas Roja
100.000–350.000	Chile habanero, Scotch Bonnet, chile dátil, Capsicum chinense
100.000–200.000	Rocoto, chile jamaicano picante, piri piri
50.000–100.000	Chile thai, chile malagueta, chile chiltepín, chile piquín
30.000–50.000	Pimienta roja o de cayena, ají escabeche, chile tabasco, algunas tipos de chile chipotle
10.000–23.000	Chile serrano, algunos tipos de chile chipotle
5.000–8.000	Variedad Nuevo México chile anaheim, chile húngaro de cera
2.500–5.000	Chile jalapeño, pimienta de Padrón, salsa tabasco
1.500–2.500	Chile rocotillo
1.000–1.500	Chile poblano
500–1.000	Chile anaheim
100–500	Pimiento, pepperoncini
0	No picante, pimiento verde

IV.4.6. Determinación de capsaicinoides en pimentones agridulces.

El contenido de C y DHC se ha determinado en pimentones agridulces por el método descrito. Los resultados obtenidos se muestran en la tabla A1.4. del anexo 1 para las muestras de pimentones agridulces analizadas mediante HPLC. Se han incluido los valores que están fuera de LOQ y LOD según recomendación de la IUPAC [83]. En este caso, hay una cierta dispersión entre los resultados obtenidos, debida a que el proceso de obtención de pimentones agridulces se basa en un mezclado manual de

pimentones dulces y picantes. En muchos casos los resultados están por debajo de los LOD y LOQ pero se han estimado las concentraciones de capsaicina y dihidrocapsacina. En otros casos encontramos valores por encima tanto de LOD como LOQ.

En la figura IV.117. se muestra un cromatograma correspondiente a una muestra de pimentón agridulce analizada mediante HPLC según las condiciones mostradas en la tabla IV.38. En la tabla IV.47. se recogen los valores de la mediana, los mínimos y los máximos para las concentraciones de C y DHC de ambos orígenes así como los del parámetro SHU.

Tabla IV.47. Valores de mediana, mínimo y máximo de la concentración (mg kg^{-1}) de C y DHC y el parámetro SHU para pimentones agridulces.

Variable	MURCIA	EXTREMADURA
	Agridulce	
C	7,5	6,2
	(2,8-14,7)	(1,3-20,1)
DHC	7,4	6,0
	(2,8-13,8)	(1,6-20,6)
SHU	121,5	99,0
	(45-237)	(21-324)

Valores mínimos y máximos entre paréntesis. C: capsaicina, DHC: dihidrocapsaicina; SHU: unidades Scoville.

En la tabla IV.47. se observa como dentro del origen Extremadura se encuentra la muestra mas picante con una concentracion de capsaicina de $20,1 \text{ mg kg}^{-1}$. El valor mínimo de capsaicina también se encuentra en una muestradel origen Extremadura, y el valor de la mediana más alto lo presentan las muestras de Murcia. Con respecto a la concentracion de dihidrocapsaicina, la situación es identica a la encontrada para la capsaicina. En algunas muestras no se ha detectado capsaicina ni dihidrocapsaicina.

La figura IV.117 muestra un cromatograma correspondiente a una muestra de pimentón agridulce. Se aprecian los picos de capsaicina y dihidrocapsaicina, que tienen menor intensidad que en el caso de los cromatogramas de muestras de pimentón picante.

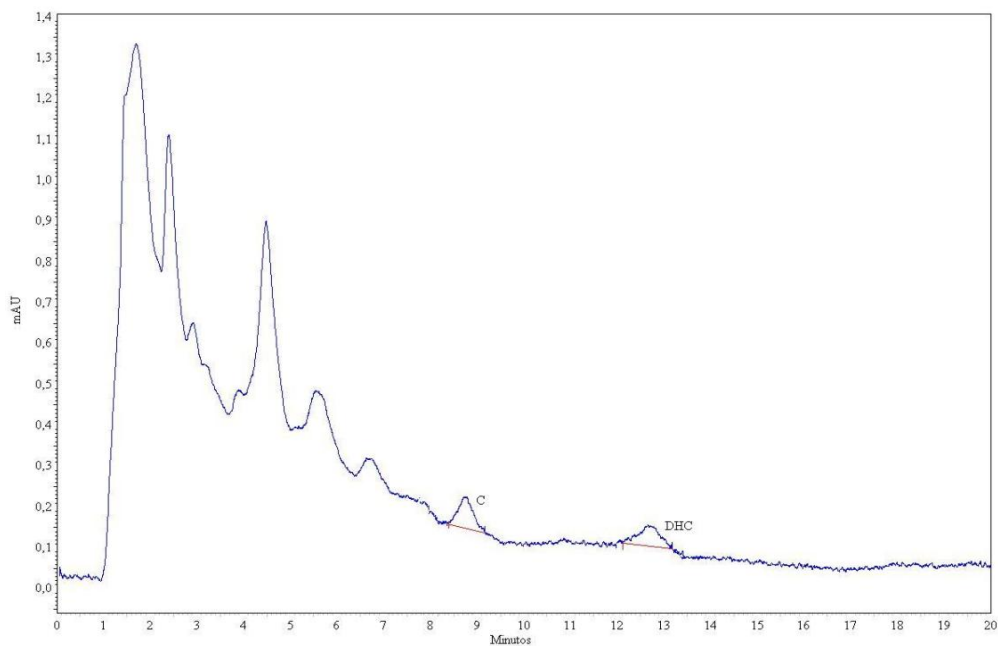


Figura IV.117. Cromatograma de una muestra de pimentón agridulce. Condiciones cromatográficas tabla IV.38.

En cuanto al parámetro SHU, los resultados obtenidos están en el rango de 0-324 unidades SHU, lo que significa que según la escala de Scoville, no tienen sabor picante o se trata de un ligero sabor picante, ya que están en la parte más baja de la escala. Los valores máximos y mínimos encontrados para los pimentones agridulces los encontramos para el origen Extremadura, aunque como se explicaba anteriormente, la dispersión de los datos es muy grande debido a la aleatoriedad del mezclado. El rango de unidades SHU es de 0-324 para las muestras de Extremadura y 0-237 SHU para el origen Murcia. Los valores de la mediana de SHU para Extremadura y Murcia son 99 y 121,5, respectivamente.

IV.4.7. Determinación de capsaicinoides en pimentones dulces.

En el caso de los pimentones dulces, se realizó el mismo procedimiento de tratamiento de la muestra y posterior análisis mediante HPLC. Los valores de concentración de capsaicina y dihidrocapsaicina fueron mucho menores que en el caso de las muestras de pimentón picante o agridulce. En muchas de las muestras no se detectaron ninguno de los dos compuestos.

En la tabla A3.5. del anexo 1 se recogen los resultados obtenidos en la determinación de capsaicinoides de pimentones dulces de los orígenes de Murcia y Extremadura. En la tabla IV.48. se muestran los valores de la mediana, mínimos y máximos para la concentración de C, de DHC y el parámetro SHU.

Tabla IV.48. Valores de mediana, mínimo y máximo de la concentración (mg kg^{-1}) de C y DHC y el parámetro SHU para pimentones dulces.

Variable	MURCIA	EXTREMADURA
	Dulce	
C	5,0	3,6
	(n.d.-8,9)	(n.d.-7,8)
DHC	2,1	2,1
	(n.d.-4,7)	(n.d.-3,4)
SHU	77,0	57,0
	(n.d.-144)	(n.d.-126)

Valores mínimos y máximos entre paréntesis. C: capsaicina, DHC: dihidrocapsaicina, SHU: unidades Scoville.

Al ser pimentones dulces, el sabor picante no es una característica propia de ellos, y se ve reflejado en la concentración, especialmente, de capsaicina. Para los pimentones del origen Extremadura, el valor máximo de C y de DHC es 7,8 y 3,4 mg kg^{-1} , respectivamente, mientras que los procedentes de Murcia presentan máximos de 8,9 y 4,7 mg kg^{-1} de C y DHC. Todos los pimentones presentan un valor de SHU en el

rango de 42-144, que como puede observarse en la tabla IV.46. se corresponde con un rango en el que los pimientos o pimentones no pican.

El valor de SHU más alto se presenta en una muestra de Murcia, mientras que el más bajo que se ha podido detectar se corresponde con el origen Extremadura, siendo estos 144 y 42, respectivamente. El rango de unidades SHU para las muestras de Extremadura es de 0-126 siendo 42 el valor mínimo que se ha podido observar y 0-144 SHU para el origen Murcia, siendo el valor más bajo de 47. Los valores de la mediana de SHU para Extremadura y Murcia son 57 y 77, respectivamente.

IV.4.8. Modelos para predecir las unidades SHU en pimentones picantes a partir del espectro de UV-vis.

En este apartado se incluye el estudio realizado para correlacionar el grado de picor de los pimentones con los espectros de absorción molecular en el UV-Vis. A partir de los espectros registrados según el procedimiento descrito en el apartado III.2.2. y las concentraciones de capsaicina, se pretende generar un modelo que permita predecir los valores SHU para muestras de pimentones picantes. La construcción de modelos apropiados que permitan predecir el valor de SHU a partir de espectros de UV-Vis, implicaría conocer estos valores mediante una técnica simple, rápida y barata. Como se ha explicado anteriormente, es posible determinar la concentración de capsaicinoides mediante técnicas como HPLC y de ahí calcular el valor de SHU. Pero este tipo de técnicas suponen un coste económico y de tiempo elevado debido principalmente al pretratamiento y limpieza de las muestras, así como el tiempo necesario para el registro de los cromatogramas y acondicionamiento del equipo. El procedimiento que se propone aquí, al tratarse del registro de un espectro UV-vis, reduce el pretratamiento de las muestras y el tiempo de análisis se acorta considerablemente. El coste económico es menor que el de un método cromatográfico, pues se reduce el uso de disolventes y se elimina el uso de precolumnas, necesarias para evitar que las muestras de pimentón dañen la columna cromatográfica. Existe otra forma de determinar de manera cualitativa la concentración de capsaicinoides, y es mediante catas ciegas de muestras de pimiento y pimentón. Esta práctica, además de ser cualitativa, tiene un coste más elevado que las

medidas espectrofotométricas, ya que para realizar una cata de pimientos y pimentones es necesario disponer de un personal altamente cualificado y un número de muestras amplio para preparar el panel de cata. El tiempo que se consume en una cata es otro inconveniente frente a las medidas de UV-vis, ya que existe en el proceso una primera fase visual, y posteriormente una olfativa y otra gustativa [88].

Existen estudios en la bibliografía sobre la predicción de parámetros cualitativos a partir de variables químicas donde se construyen modelos mediante métodos de regresión lineal y no lineal. Sousa et al. [89,90] proponen modelos de predicción de la concentración de ozono por hora en el aire a partir de MLR y ANN y datos provenientes de PCA. Las variables predictoras empleadas son la concentración en el aire de NO, NO₂ y datos meteorológicos como la velocidad del viento, la humedad relativa y la temperatura ambiental. Chauchard et al. [91] determinaron la acidez total en uvas mediante un modelo creado a partir de datos de NIR y PLS. Mo et al. [92] determinaron la concentración de capsaicinoides en pimientos rojos a partir de un modelo creado por PLSR y usando como variables predictoras datos de espectroscopia UV-Vis-NIR.

Generalmente se suelen usar técnicas de regresión lineal en la construcción de estos modelos, pero se pueden presentar dos problemas: multicolinealidad y alta dimensionalidad de las variables predictoras. En este estudio, las variables están muy correlacionadas entre sí, debido a que son valores de absorbancia de un espectro continuo como el de UV-vis. Además existe una alta dimensionalidad puesto que existe una variable, el valor de absorbancia, para cada longitud de onda. Las soluciones para estos problemas son la aplicación de PLS o PCA. En ambos casos, las variables predictoras originales se transforman en nuevas variables ortogonales que corrigen el problema de la multicolinealidad y reducen el espacio de las variables predictoras.

En este caso, se han empleado técnicas lineales como regresión lineal múltiple, PLS, PCA y técnicas no lineales como ANN para la construcción de los modelos. Se proponen cuatro modelos, evaluando sus resultados en busca de la mejor solución al problema.

Tabla IV.49. Rangos codificados SHU.

5.000–8.000	G
2.500–5.000	F
1.500–2.500	E
1.000–1.500	D
500–1.000	C
100–500	B
0	A

Se pretende obtener un modelo que permita predecir el rango de SHU al que pertenecen las muestras de pimentón picante. Para evaluar los modelos se estudian los resultados en términos del intervalo de unidades SHU al que pertenecen las muestras. Para ello, los intervalos se codifican tal y como se muestra en la tabla IV.49, y los resultados obtenidos individualmente por el modelo para cada muestra se transforman en los códigos correspondientes a dicho intervalo. Por ejemplo, si se determina experimentalmente que la muestra presenta un valor SHU dentro de un intervalo de la tabla IV.49. y el valor predicho por el modelo pertenece al mismo intervalo se considera un acierto, en caso contrario se considera un error.

Las muestras elegidas para este estudio son las que pertenecen a los intervalos E (n=17) y F (n=10). En el caso de A, B, C, D y G no existe un número de muestras representativo para realizar un cálculo adecuado de los modelos.

Para hacer una validación cruzada de los modelos de clasificación se aplica, DAGKJ, creándose nueve replicados en la matriz de las 27 muestras. Para los conjuntos de entrenamiento se emplean 2/3 de las muestras y para el conjunto prueba 1/3 del total. En los casos en los que se aplican ANN, el conjunto de entrenamiento se subdivide en un 50% validación y 50% entrenamiento.

Se han estudiado diferentes modelos con el objetivo de conocer cuál proporciona mejores resultados de predicción a la hora de conocer el grado de picor de pimientos y pimentones, a partir de datos obtenidos mediante HPLC y transformados en unidades SHU. Todos los modelos parten de los espectros UV-Vis correspondientes a las muestras de pimentón picante. El rango de longitud de onda que se emplea para este estudio va desde los 215 a los 315 nm, ya que existen dos picos característicos para los

capsaicinoides dentro de ese rango a 230 y 280 nm [93]. Varios autores utilizan, también, un rango de longitud de onda similar, ya que los espectros no introducen ruido debido a que la mayoría de los carotenoides absorben a longitudes de onda mayores, entre 560 y 580 nm, principalmente [94,95]. Además, en un estudio realizado por Davis et al. [96] sobre capsaicinoides, registraron un espectro en el rango de 190 a 1100 nm y seleccionaron la región de 215 a 300 nm para los estudios de los modelos regresión. Como ya se ha comentado previamente, en este caso existen problemas de multicolinealidad y alta dimensionalidad por lo que es necesario recurrir a otras variables diferentes de las originales. Para seleccionar el número de variables transformadas que se emplearan en los modelos, es necesario un criterio de selección. El criterio empleado para retener los PCs es el criterio $PRESS_{f+1}/PRESS_f > 1$, donde PRESS se calcula como:

$$PRESS_f = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.72)$$

siendo \hat{y}_i el valor predicho de la *i*-ésima respuesta observada cuando se deja un conjunto fuera del entrenamiento e y_i el valor real de la *i*-ésima respuesta observada. Según este criterio, se siguen incluyendo PCs hasta que el cociente es mayor que 1, lo que indicaría que la inclusión de un nuevo PC no favorece los resultados del modelo.

El error PRESS, que tiene en cuenta los valores individuales, da una idea de la calidad del modelo, pero éste se evalúa teniendo en consideración cuántas muestras dentro de cada conjunto del DAGJK se clasifican correctamente en el intervalo al que pertenecen. La eficacia del modelo se calcula a partir del cociente de muestras correctamente clasificadas entre muestras totales. La idea de evaluar los modelos mediante intervalos de la escala SHU se basa, principalmente, en que la conversión de la concentración de capsaicinoides a unidades de parámetros cualitativos y viceversa puede introducir errores de más del 20%, como ocurre en el caso del parámetro ASTA [97], y hay que recordar que la escala SHU se trata de una escala organoléptica creada a partir de muestras naturales [85].

- Modelo I

El modelo I se basa en la aplicación de PCA al conjunto de datos, seguido de una regresión lineal múltiple. Cuando estas dos herramientas se combinan hablamos de PCR. Se trata de una alternativa para poder tratar datos correlacionados entre sí, que solamente con regresión lineal múltiple no se podrían tratar.

En PCR, tal y como se procede en un modelo lineal múltiple, se establece una relación funcional entre variables dependientes Y e independientes X. Esta relación permitirá realizar predicciones de las variables del conjunto Y mediante las proyecciones ortogonales de las columnas de esa matriz, sobre el espacio generado por las columnas de X.

Inicialmente se hace un cálculo de PCs a partir de los valores de absorbancia correspondientes al intervalo de longitud de onda de 215 a 315 nm. Se obtienen los PCs que expliquen la mayor varianza posible en los datos originales y usando estos PCs como variables se construye el modelo de regresión.

En el caso del modelo I, se construye con los PCs con autovalores mayores que la unidad y se van incluyendo según el criterio previamente mencionado. En la siguiente tabla (IV.50.) se puede observar los autovalores y la varianza explicada por los cinco primeros PCs.

Tabla IV.50. Autovalores, varianza y varianza acumulada para pimentones picantes en el modelo I.

Nº de PCs	Autovalor	Varianza explicada	Varianza total
PC1	99,64	98,65	98,65
PC2	1,09	1,08	99,73
PC3	0,15	0,15	99,88
PC4	0,07	0,07	99,94
PC5	0,04	0,04	99,98

La primera variante del modelo I, se construye con los dos primeros PCs, ya que son los que presentan autovalores mayores que uno. Se van añadiendo PCs a los

cálculos hasta que el cociente $PRESS_{f+1}/PRESS_f$ sea mayor que uno. Los errores PRESS calculados se presentan en la tabla IV.51.

Tabla IV.51. Errores PRESS para los cálculos con distinto número de PCs.

Nº de PCs	PRESS*	$PRESS_{f+1}/PRESS_f$
2	9,36	
3	10,12	1,0810574
4	9,60	0,9491521
5	10,71	1,1156004

* $PRESS \times 10^{-6}$.

Como se puede observar en la tabla IV.51., cuando el modelo se construye con los dos primeros PCs, se obtienen mejores resultados que al añadir un tercer PC. Sin embargo, como el valor obtenido no es mucho mayor que la unidad, se siguen añadiendo más PCs para comprobar que efectivamente no existen mejoras importantes. Además, no es aconsejable calcular modelos con tan pocas variables [98]. En el caso de la inclusión de un cuarto PC, se obtiene un error PRESS menor, y por tanto un resultado mejor que el modelo con 3PCs. Al incluir un quinto PC, PRESS se hace mayor y el cociente de la ecuación 4.72. es mayor que uno.

En la tabla IV.52.se muestran los porcentajes de eficacia para cada variante del modelo I, así como la desviación estándar a partir de los replicados DAGJK. En el caso del modelo calculado a partir 4 PCs, se observa mayor eficacia que en los demás casos.

Tabla IV.52. Eficacia del modelo I según nº de PCs.

Nº de PCs	Eficacia (%)
2	41 ± 6
3	42 ± 12
4	48 ± 11
5	46 ± 15

La inclusión de un cuarto PC mejora la eficacia del modelo con respecto al caso de 2,3 y 5 PCs. Sin embargo, los resultados obtenidos no son suficientemente buenos y por ello se recurre a técnicas de clasificación no lineales, ya que el problema de predecir los intervalos SHU es complejo y puede que las técnicas lineales no sean una herramienta adecuada para este caso.

▪ Modelo II:

En este segundo modelo se ha considerado una aproximación no lineal construyendo una ANN. Los scores de los PCs se emplean como variables independientes y como variable dependiente los valores SHU. Inicialmente se construye una red con los dos primeros componentes principales y el número de PCs incluidos se decide con el mismo criterio de minimizar la relación de PRESS previamente propuesta. La tabla IV.53. incluye los resultados obtenidos.

Tabla IV.53. Errores PRESS para los cálculos del modelo II con distinto número de PCs.

Nº de PCs	PRESS*	PRESS _{f+1} /PRESS _f
2	8,737	
3	6,203	0,7099647
4	6,798	1,0958567
5	5,702	0,8388434

*PRESSx10⁻⁶.

Para los modelos con 2, 3,4 y 5 PCs, siguiendo el criterio PRESS_{f+1}/PRESS_f la inclusión de un tercer PC sería suficiente, ya que el error PRESS es menor en el caso de 3PCs. Sin embargo se han realizado modelos con 4 y 5 PCs, para comprobar los valores de la eficacia. En todos los casos se han empleado redes neuronales del tipo MLP-ANN. En el primer caso, se obtiene una red neuronal de arquitectura 2:5:1. Las funciones de activación de la capa oculta y de la de salida son exponencial y tangencial, respectivamente. Se obtiene un 46% de eficacia, lo cual no mejora los resultados obtenidos en el modelo I. En el caso de las MLP-ANN, es conveniente emplear más de dos variables para el cálculo [99]. Los resultados de eficacia se recogen en la tabla IV.54

Tabla IV.54. Eficacia del modelo II según n° de PCs.

n° de PCs	Eficacia (%)
2	54 ± 18
3	67 ± 23
4	64 ± 22
5	68 ± 22

Se construye un segundo modelo empleando un PC más con una arquitectura 3-4-1 y la función de activación tangencial tanto para la capa oculta como para la de salida. Se obtiene un menor valor de PRESS y un mayor porcentaje de acierto en este caso que cuando se obtiene el modelo con 2 PCs. Si se incluye un cuarto PC, la mejor red obtenida es de arquitectura 4:10:1 con función de activación exponencial para la capa oculta y tangencial para la de salida. Esta nueva versión del modelo II lleva a una situación en la que $PRESS_{4PCs} > PRESS_{3PCs}$ y la eficacia es menor. Si se incluye un quinto PC, al igual que se hizo en el modelo I, nos encontramos con un error PRESS menor que en cualquier caso, además de un porcentaje de eficacia mayor. Para todos estos modelos la desviación estándar obtenida a partir del DAGJK es comparable con independencia del número de componentes incluidos.

▪ **Modelo III**

Este modelo se puede considerar como una generalización de la MRL, con la ventaja de que al incluir un análisis PLS se pueden analizar datos fuertemente correlacionados entre sí o con ruido debido a variables que no proporcionan ninguna información a los problemas de predicción o clasificación. PLS es una alternativa muy eficiente pero poco usada en quimiometría según Frank y Friedman [100]. Con este modelo, se consiguen nuevas variables que son estimaciones de las variables latentes y se denominan t-scores. Estas variables nuevas, t-scores, son las que se usan como variables independientes (predictoras) en el modelo de regresión. Al igual que ocurría con los PCs, es necesario saber qué componentes son los que explican mayor varianza y tienen, por tanto, autovalores mayores. En la tabla IV.55. se recogen estos resultados. En este caso, el primer componente es el único que tiene un autovalor mayor que uno, aunque se han calculado los modelos con 2, 3, 4, y 5 t-scores, evaluando el error PRESS para decidir qué número de variables incluir en el modelo III.

Tabla IV.55. Autovalores, varianza y varianza acumulada para pimentones picantes en el modelo III.

Nº de t-scores	Autovalor	% varianza explicada	% varianza total
t-scores 1	92,45	91,53	91,53
t-scores 2	0,71	0,70	92,23
t-scores 3	0,21	0,20	92,44
t-scores 4	0,04	0,04	92,47
t-scores 5	0,01	0,01	92,49

Como se ha indicado, los t-scores de las muestras se emplean como variables independientes, x_i , mientras que los valores de SHU son las variables dependientes, y_i . Los valores SHU se predicen mediante MLR, calculando los valores PRESS (tabla IV.56.) y obteniendo una eficacia de predicción a partir de los valores SHU de las muestras del conjunto test para nueve replicados, según el procedimiento DAGKJ (tabla IV.57.).

Tabla IV.56. Errores PRESS para los cálculos del modelo III con distinto número de t-scores.

Nº de t-scores	PRESS*	$PRESS_{f+1}/PRESS_f$
2	10,264	
3	9,315	0,9074977
4	7,272	0,7806889
5	8,782	1,2076285

*PRESS $\times 10^{-6}$.

Al igual que ocurría con el modelo I, los mejores resultados de PRESS se obtienen cuando se emplean 4 variables de entrada, t-scores en este caso. La inclusión de un quinto t-score aumenta el valor de PRESS, con lo que es de suponer que la eficacia del modelo disminuya respecto a la variante con 4 t-scores.

Tabla IV.57. Eficacia del modelo III según nº de t-scores.

Nº de t-scores	Eficacia (%)
2	50 ± 11
3	67 ± 16
4	70 ± 16
5	69 ± 17

El porcentaje de acierto en la predicción de los intervalos con 2 t-scores es de un 50% con una desviación estándar de 11. Cuando se hace el cálculo con más t-scores la eficacia aumenta hasta un 70 %. Cuando se utilizan más variables puede ocurrir que las nuevas variables incluyan más ruido sin mejora de la eficacia [101], pero en este caso la desviación estándar de las variantes del modelo II son similares.

Los modelos III, donde se emplean exclusivamente técnicas lineales, permiten obtener resultados que mejoran notablemente respecto a PCR (modelo I).

▪ Modelo IV

En este modelo, como alternativa no lineal, se utilizan MLP- ANNs para llevar a cabo la regresión. La matriz de datos se compone de los t-scores obtenidas por PLS como variables independientes, y los valores SHU como variables dependientes.

Se construyen sendas redes con los 2 y 3 t-scores de arquitectura 2:7:1 y 3:5:1, respectivamente. La función de la capa oculta en ambos casos es tangencial, mientras que la función de la capa de salida es logística y exponencial, respectivamente. La tabla IV.58 muestra los valores de PRESS obtenidos. Cuando se emplean tres t-scores el error PRESS disminuye. El cálculo se repite considerando cuatro y cinco t-scores, para comprobar posteriormente la eficacia del modelo. Al añadir un t-score más se obtiene un error PRESS menor que en cualquiera de los demás modelos, por lo tanto, es presumible que el valor de eficacia también será mayor en este caso. Se añade un t-score más para comprobar que no sigue disminuyendo el error PRESS.

Tabla IV.58. Errores PRESS para los cálculos del modelo IV con distinto número de t-scores.

Nº de t-scores	PRESS*	PRESS _{f+1} /PRESS _f
2	7,869	
3	5,911	0,7511619
4	4,144	0,7010083
5	5,538	1,3364193

*PRESSx10⁻⁶.

Los valores de eficacia junto con la desviación estándar obtenidos se muestran en la tabla IV.59. La eficacia conseguida con el modelo IV con 4 t-scores es 80%. Este valor es el más alto de todos los modelos obtenidos.

Tabla IV.59. Eficacia del modelo IV según nº de t-scores.

Nº de t-scores	Eficacia (%)
2	58 ± 17
3	71 ± 24
4	80 ± 13
5	61 ± 17

En la figura IV.118 se comparan los resultados obtenidos, relativos a su eficacia, para cada modelo junto con el error asociado a los replicados del DAGJK. Como puede observarse, los mejores resultados se tienen para el modelo PLS-MLP-ANN (modelo IV) empleando las cuatro primeras variables latentes como datos de entrada.

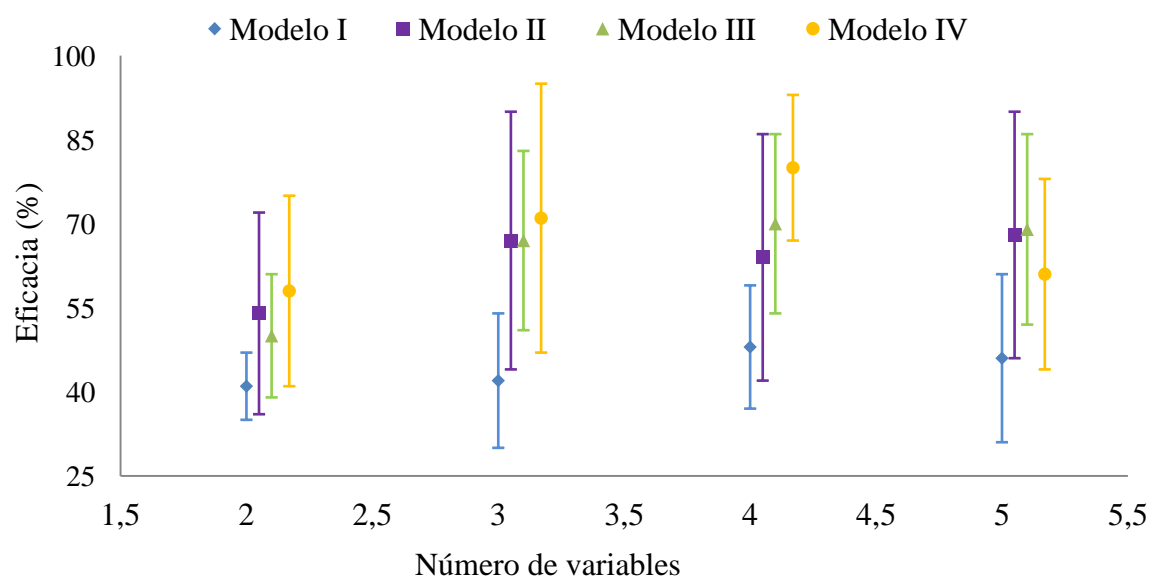


Figura IV.118. Eficacia de los distintos modelos.

IV.5. Bibliografía

1. A. G. González, M. A. Herrador, A. G. Asuero. Practical digest for evaluating the uncertainty of analytical assays from validation data according to the LGC/VAM protocol. *Talanta*, **2005**, 65, 1022–1030.
2. V. J. Barwick, L. R. Ellison, *VAM Project 3.2.1, Development and Harmonisation of Measurement Uncertainty Principles. Part D: Protocol for Uncertainty Evaluation from Validation Data*, January, 2000, Report No.: LGC/VAM/1998/088.
3. A. G. González, M. A. Herrador, A. G. Asuero. Intra-laboratory testing of method accuracy from recovery assays. *Talanta*, **1999**, 48, 729-736.
4. J. A. Rafter, M. L. Abell, J. P. Braselton. Multiple comparison methods for means. *SIAM Review*, **2002**, 44, 259-278.
5. EURACHEM/ CITAC Guide, *Quantifying Uncertainty in Analytical Measurement*, 2nd., 2000.
6. R. Compañó, A. Ríos. *Garantía de la Calidad en los Laboratorios Analíticos*. Síntesis, Madrid, 2002.
7. Y. P. Kalra (Ed). *Handbook of Reference Methods for Plant Analysis Council*, CRC Press, Boca Ratón, 1998.
8. B. D. Hall, R. Willink. Does "Welch-Satterthwaite" make a good uncertainty estimate? *Metrologia*, **2001**, 38, 9-15.
9. International Organization of Standardization (ISO), ISO/DTS 21748, *Guide to Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation*, ISO, Geneva, Switzerland, 2003.
10. R. Wood. How to validate analytical methods. *Trends in Analytical Chemistry*, **1999**, 18, 624-632.
11. W. Horwitz, R. Albert. The Horwitz ratio (HorRat): A useful index of method performance with respect to precision. *Journal of AOAC International*, **2006**, 894, 1095-1109.
12. L. Cuadros, A. M. García, J. M. Bosque. Statistical estimation of linear calibration range. *Analytical Letters*, **1996**, 29, 1231-1239.
13. J. Inczedy, T. Lengyel, A. M. Ure. *Compendium of Analytical Nomenclature (definitive rules 1997) 3rd.*, Blackwell Science, Oxford, 1998.

14. J. Mataix Verdú, *Tabla de Composición de Alimentos*, Universidad de Granada, 2009.
15. M. Brunner, R. Katona, Z. Stefánka, T. Prohaska. Determination of the geographical origin of processed spice using multielement and isotopic pattern on the example of Szegedi paprika. *European Food and Research Technology*, **2010**, 231, 623–634.
16. A. V. Filgueiras, J. L. Capelo, I. Lavilla, C. Bendich. Comparison of ultrasound-assisted extraction and microwave-assisted digestion for determination of magnesium, manganese and zinc in plant samples by flame atomic absorption spectrometry. *Talanta*, **2000**, 53, 433–444.
17. A. Rial Boubeta, *Estadística Práctica para la Investigación en Ciencias de la salud*. Netbiblo, La Coruña, 2008.
18. J. Muth. *Basic Statistics and Pharmaceutical Statistical Applications*. Marcel Dekker, New York, 1999.
19. S. Martín Fernández. *Guía Completa de Statgraphics; desde MS-DOS a Statgraphics Plus*. Díaz de Santos, Madrid, 2001.
20. I. T. Jolliffe. *Principal Components Analysis*, Springer, New York, 2002.
21. W. P. Gardiner, *Statistical Analysis Methods for Chemistry*, Royal Society of Chemistry, Cambridge, 1997.
22. S. Zomer, R. G. Brereton, J. F. Carter, C. Eckers. Support vector machines for the discrimination of analytical chemical data: application to the determination of tablet production by pyrolysis-gas chromatography-mass spectrometry. *Analyst*, **2004**, 129, 175-181.
23. D. L. Massart. *Handbook of Chemometrics and Qualimetrics- Part B*, Elsevier, Amsterdam, 1998.
24. W. S. Sarle. *Neural Networks and Statistical Models*, Ed. Proceedings of the 19th Annual SAS Users Group International Conference, 1994, 1538–1550.
25. P. S. Kott. The Delete-a-Group Jackknife. *Journal of Official Statistics*, **2001**, 17, 521–526.
26. M. Forina, C. Armanino, R. Leardi, G. Drava, J. A class-modelling technique based on potential functions. *Journal of Chemometrics*, **1991**, 5, 435-453.
27. H. F. Kaiser, The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, **1960**, 20, 141-151.

28. S. Tsakovski, V. Simeonov. *Chemometrics as a Tool for Treatment Processing of Multiparametric Analytical Data Set*, Ed. J. Namiesnik, P. Szefer, *Analytical Measurements in Aquatic Environments*, CRC, Boca Ratón, 2009.
29. A. Palacios-Morillo, A. Alcázar, F. Pablos, J. M. Jurado. Differentiation of tea varieties using UV–Vis spectra and pattern recognition techniques. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **2013**, 103, 79-83.
30. C. Lau, C. Chan, F. Chau, D. K. Mok. Rapid analysis of radix puerariae by near-infrared spectroscopy. *Journal of Chromatography A*, **2009**, 1216, 2130-2135.
31. D. L. B. Wetzel. *Instrumental Methods in Food and Beverage Analysis*. Elsevier, Amsterdam, 1998.
32. OIV. *Determination of Chromation Characteristics According to CIELab (Method OIV-MA_AS2-11)*. Compendium of International Analysis of Methods-OIV, 2006.
33. CIE Standard Colorimetric observers <http://www.cis.rit.edu/mcsl/online/cie.php>.
34. A. Pérez Gálvez, M. I. Mínguez Mosquera, J. Garrido Fernández, M. M. Lozano Ruíz, V. Montero de Espinosa. Correlation between ASTA units-carotenoid concentration in paprika. Prediction of the color stability during storage. *Grasas y Aceites*, **2004**, 55, 213-218.
35. E. J. Gilabert. *Medida del Color*, Ed., U.P.V. Escuela politécnica de Valencia. Valencia, 1998.
36. Commission Internationale de l'Eclairage. *Technical report CIE 15.2004, Colorimetry 3rd.*, 2004.
37. J. A. González, E. Perino, M. C. Ruíz. Blanqueado de arcillas mediante la eliminación de hierro en la etapa de quemado. *Jornadas SAM/CONAMET/SIMPOSIO MATERIA*, San Luis, Argentina, 2010.
38. BOE 05-01-2006, núm 4, P. 778, Ministerio de Agricultura, Pesca y Alimentación.
39. Reglamento (CE) núm. 464/2001 07-03-2001. (DOCE 8-03-2001, L. 66, P. 0029-0030).
40. A. Topuz, H. Feng, M. Kushad. The effect of drying method and storage on color characteristics of paprika. *LWT - Food Science and Technology*, **2009**, 42, 1667–1673.
41. D. Giuffrida, P. Dugo, G. Torre, C. Bignardi, A. Cavazza, C. Corradini, G. Dugo. Evaluation of carotenoid and capsaicinoid contents in powder of red chili peppers during one year of storage. *Food Research International*, **2014**, 65B, 163-170.

42. D. R. Tainter, A. T. Grenis. *Spices and Seasonings: A Food Technology Handbook*, 2nd., Willey-VCH, New York, 2001.
43. J. M. Nieto-Sandoval, J. A. Fernández-López, L. Almela, J. A. Muñoz. Dependence between apparent color and extractable color in paprika. *Color. Research and Application*, **1999**, 24, 93–97.
44. M. Krajayklang, A. Klieber, P. R. Dry. Colour at harvest and post-harvest behavior influence paprika and chilli spice quality. *Postharvest Biology and Technology*, **2000**, 20, 269-278.
45. F. E. Quipo Muñoz, A. M. Ramírez Muñoz, J. A. Rojas Pérez, L. E. Ordoñez Santos. Cambios en la vitamina C y el color durante la cocción del pimentón verde (*Capsicum Annuum* L). *Tecno Lógicas*, **2013**, 31, 141-150.
46. J. V. Collantes Duarte. *Predicción con Redes Neuronales: Comparación con las Metodologías de Box y Jenkins*. Magister Scientiae en Estadística Aplicada. Universidad de los Andes. Mérida, Venezuela. 2001, 25-45.
47. D. Ortiz, F. Vila, J. Velásquez. *Revista Avances en Sistemas e Informática*, **2007**, 4, 134-144.
48. M. T. Hassan Than. *Recent Trends on QSAR in the Pharmaceutival Perceptions*. Bentham Science Publishers, Amsterdam, 2012.
49. J. M. Marín, *Análisis de componentes principales*. Material docente. <http://halweb.uc3m.es/esp/personal/personas/jmmarin/esp/AMult/tema3am.pdf>
50. C. J. C. Burges. *A tutorial on Support Vector Machines for Pattern Recognition*, 1998 (<http://svms.org/tutorials/Burges1998.pdf>).
51. R. Karnkca, M. Rayanakorn, S. Watanesk, Y. Vaneesorn. Optimization of high-performance liquid chromatographic parameters for the determination of capsaicinoid compounds using the simplex method. *Analytical Sciences*, **2002**, 18, 661-665.
52. A. Laskaridou-Monnerville. Determination of capsaicin and dihydrocapsaicin by micellar electrokinetic capillary chromatography and its application to various species of *Capsicum*, Solanaceae. *Journal of Chromatography A*, **1999**, 838, 293-302.
53. L. Liu, X. Chen, L. Juli, X. Deng, W. Duan, S. Tan. Determination of capsaicin and dihydrocapsaicin in *Capsicum annuum* and related products by capillary electrophoresis with a mixed surfactant system. *Food Chemistry*, **2010**, 119, 1228-1232.

54. J. Juangsamoot, C. Ruangviriyachai, S. Techawongstien, S. Chanthai. Determination of capsaicin and dihydrocapsaicin in some hot chilli varieties by RP-HPLC-PDA after magnetic stirring extraction and clean up with C18cartridge. *Journal of International Food Research*, **2012**, 19, 1217-1226.
55. H. G. Daood, V. Illés, M. H. Gnayfeed, B. Mészáros, G. Horvth, P. A. Biacs. Extraction of pungent spice paprika by supercritical carbon dioxide and subcritical propane. *Journal of Supercritical Fluids*, **2002**, 23, 143-152.
56. G. F. Barbero, M. Palma, C. G. Barroso. Determination of capsaicinoids in peppers by microwave-assisted extraction-high-performance liquid chromatography with fluorescence detection. *Analytica Chimica Acta*, **2006**, 578, 227-233.
57. P. Kirschbaum-Titze, C. Hiepler, E. Mueller-Seitz, M. Petz. Pungency in Paprika (*Capsicum annuum*). Decrease of capsaicinoid content following cellular disruption. *Journal of Agricultural and Food Chemistry*, **2002**, 50, 1260-1263.
58. E.S. Poyrazoglu, O. Yemis, C. Kadacal, N. Artic. Determination of capsaicinoid profile of different chilli peppers grown in Turkey. *Journal of the Science of Food and Agriculture*, **2005**, 8, 1435-1438.
59. B. Estrada, F. Pomar, J. Díaz, F. Merino, M. A. Bernal. Pungency level in fruits of the Padron pepper with different water supply. *Scientia Horticulturae*, **1999**, 81, 385-396.
60. W. S. Hawer, J. H. Ha, J. B. Hwang, Y. J. Nam. Effective separation and quantitative-analysis of major heat principles in red-pepper by capillary gas-chromatography. *Food Chemistry*, **1994**, 49, 99-103.
61. M. S. Chinn, R. R. Shama-Shiyappa, L. Jacqueline, L. Cotter. Solvent extraction and quantification of capsaicinoids from *Capsicum chinense*. *Foods and Bioproducts Processing*, **2011**, 89, 340-345.
62. B. J. Rowland, B. Villalon, E. E. Burns. Capsaicin production in sweet bell and pungent Jalapeño peppers. *Journal of Agricultural and Food Chemistry*, **1983**, 31, 484-487.
63. B. Kauffmann, P. Chirsten. Recent extraction techniques for natural products: microwave-assisted extraction and pressurized solvent extraction. *Phytochemical Analysis*, **2002**, 13, 105-113.

64. M. D. Collins, L. M. Wasmund, P. W. Bosland. Improved method for quantifying Capsaicinoids in capsicum using high performance liquid chromatography. *Horticultural Science*, **1995**, 30, 137–139.
65. Z. A. Al Othman, Y. B. Hadj Ahmed, M. Abdelaty Habila, A. A. Ghafar. Determination of capsaicin and dihydrocapsaicin in *Capsicum* fruit samples using high performance liquid chromatography. *Molecules*, **2011**, 16, 8919-8929.
66. P. G. Hoffman, M. C. Lego, W. G. Galetto. Separation and quantitation of red pepper major heat principles by reverse-phase high-pressure liquid chromatography. *Journal of Agricultural and Food Chemistry*, **1983**, 31, 1326-1330.
67. A. Saria, F. Lembeck, G. Skofitsch. Determination of capsaicin in tissues and separation of capsaicin analogues by HPLC. *Journal of Chromatography*, **1981**, 208, 41-46.
68. D. Parriott. *A practical Guide to HPLC detection*. Academic Press, New York, 2012.
69. L. R. Snyder, J. J. Kirkland, J. L. Glajch. *Practical HPLC method development, 2nd*. John Wiley and Sons, Inc, New York, 1997.
70. J. P. Walter. Physics of high-resolution ultrasound - Practical aspects. *Radiologic Clinics of North America*, **1985**, 23, 3-11.
71. C. H. Marvin, L. Allan, B. E. McCarry, D. W. Bryant. A Comparison of ultrasonic extraction and Soxhlet extraction of polycyclic aromatic hydrocarbons from sediments and air particulate material. *International Journal of Environmental Analytical Chemistry*, **1992**, 49, 221-230.
72. A. V. Filgueiras, J. L. Capelo, I. Lavilla, C. Bendicho. Comparison of ultrasound-assisted extraction and microwave assisted digestion for determination of magnesium, manganese and zinc in plant samples by flame atomic absorption spectrometry. *Talanta*, **2000**, 53, 433-441.
73. A. Pastor, E. Vázquez, R. Ciscar, M. De la Guardia. Efficiency of the microwave-assisted extraction of hydrocarbons and pesticides from sediments. *Analytica Chimica Acta*, **1997**, 344, 241-249.
74. M. Contreras-Padilla. E. M. Yahia. Changes in capsaicinoids during development, maturation and senescence of chile peppers and relation with peroxidase activity. *Journal of Agricultural and Food Chemistry*, **1998**, 46, 2075-2079.

75. B. Estrada, M. A. Bernal, J. Díaz, F. Pomar, F. Merino. Fruit development in *capsicum annumm*: changes in capsaicin lignin, free phenolics and peroxidase patterns. *Journal of Agricultural Food Chemistry*, **2000**, 48, 6234-6239.
76. M. A. Rostagno, M. Palma, C. G: Barroso. Ultrasound-assisted extraction of soy isoflavones. *Journal of Chromatography A*, **2003**, 1012, 119-128.
77. Agua noG. F. Barbero, A. Liazid, M. Palma, C.G. Barroso. Ultrasound-assisted extraction of capsaicinoids from peppers. *Talanta*, **2008**, 1332-1337.
78. X. Dong, X. Li, L. Ding, F. Cui, Z. Tang, Z. Liu. Stage extraction of capsaicinoids and red pigments from fresh red pepper (*Capsicum*) fruits with ethanol as solvent. *LWT - Food Science and Technology*, **2014**, 59, 396-402.
79. D. E. Henderson, S. K. Henderson. Thermal decomposition of capsaicin. 1. Interactions with oleic acid at high temperatures. *Journal of Agricultural and Food Chemistry*, **1992**, 40,2263-2268.
80. V. K. Attuquayefio, K. A. Buckle. Rapid sample preparation method for HPLC analysis of capsaicinoids in capsicum fruits and oleoresins. *Journal of Agricultural and Food Chemistry*, **1987**, 35, 777-779.
81. A. De Castro Ríos. *Aplicación de la cromatografía líquida acoplada a la espectrometría de masas en tándem a la determinación de antidepresivos y fluido oral*. Universidad Santiago de Compostela, 2004.
82. AOAC International. *Method Validation Program (OMA/PVM Department), Including Appendix D: Guidance for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis*, 2000.
83. L. Huber. Validation of Analytical methods: review and strategy. *LC-GC International*, **1998**, 11, 96-105.
84. J. Inczèdy, T. Lengyel, A. M. Ure. *Compendium of Analytical Nomenclature. Definitive Rules. Section 18.4.3.7. Detection and Quantification Capabilities*, 3rd., 1997, IUPAC, www.iupac.org.
85. W. L. Scoville. Note on capsicums. *Journal of the American Pharmaceutical Association*, **1912**, 1, 453-454.
86. F. Hart, H. Fisher. *Análisis Moderno de Alimentos*. Acribia, Zaragoza, 1971, 410-411.

87. I. Perucka, W. Oleszek. Extraction and determination of capsaicinoids in fruit of hot pepper *Capsicum annuum* L. by spectrophotometry and high-performance liquid chromatography. *Food Chemistry*, **2000**, 71, 287-291.
88. J. López, J. Simal, L. Vázquez. Determinación de capsaicina en pimentones mediante la segunda derivada del espectro UV. *Anales de Bromatología XXXIX-1* **1987**, 161-164.
89. B.O.C. y L. n° 219, 12/11/2002, 14879, Orden de 4 de noviembre de 2002, Consejería de Agricultura y Ganadería de la Comunidad de Castilla y León, Reglamento de la indicación geográfica protegida Pimiento Asado del Bierzo. <http://www.pimientoasadodelbierzo.es/Reglamento%20IGP%20Pimiento%20del%20Bierzo.pdf>.
90. S. I. V. Sousa, F. G. Martins, M. C. M. Alvim-Ferraz, M. C. Pereira. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, **2007**, 22, 97-103.
91. S. I. V. Sousa, F. G. Martins, M. C. Pereira, M. C. M. Alvim-Ferraz. Prediction of ozone concentrations in Oporto city with statistical approaches. *Chemosphere*, **2006**, 64, 1141-1149.
92. F. Chauchard, R. Codgill, R. Roussel, J. M. Roger, V. Bellon-Maurel. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems*, **2004**, 71, 141-150.
93. C. Mo, K. Lee, J. G. Lim, S. Kang, H. D. Lee, B. K. Cho. *Development of the Pungency Measuring System for Red-pepper Powder. Sensing for Agriculture and Food Quality and Safety III*, Proc. SPIE 8027, Junio 2011.
94. U. Schweiggert, R. Carle, A. Schieber. Characterization of major and minor capsaicinoids and related compounds in chili pods (*Capsicum frutescens* L.) by high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Analytica Chimica Acta*, **2006**, 557, 236-244.
95. Y. M. A. Naguib. Antioxidant activities of astaxanthin and related carotenoids. *Journal of Agricultural and Food Chemistry*, **2000**, 48, 1150-1154.

96. B. Röhl. Carotenoid and retinoid-two pigments in a gecko eye lens. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **2000**, 125, 105-112.
97. C. B. Davis, C. E. Markey, M. A. Busch, K. W. Busch. Determination of capsaicinoids in Habanero Peppers by chemometric analysis of UV spectral data. *Journal of Agricultural Food Chemistry*, **2007**, 55, 5925-5933.
98. D. R. Tainter, A. T. Grenis. *Spices and Seasonings*, Wiley-IEEE, New York, 2001.
99. R. Boqué, J. Ferré. Análisis de componentes principales aplicado a la representación de datos multidimensionales. *Técnicas de Laboratorio*, **2004**, **290**, 214-219.
100. N. Kwak, C. H. Choi. Input feature selection for classification problems. *IEEE Transactions on Neural Networks*, **2002**, 13, 143-159.
101. I. E. Frank, J. H. Friedman. A statistical view of some chemometrics regression tools. *Technometrics*, **1993**, 35, 109-148.
102. I. T. Jolliffe. A note in the use of principal analysis in regression. *Journal of the Royal Statistical Society*, **1982**, 31, 300-303.
103. S. Burke. Missing values, outliers, robust statistics & non-parametric methods. *LC·GC Europe Online Supplement*, **1999**, 19-24.
104. S. L. R. Ellison, T. J. Farrant, V. Barwick. *Practical Statistics for the Analytical Scientist: A Bench Guide*, Royal Society of Chemistry, London, **2009**

V. Resumen y conclusiones

1. Se ha estudiado el perfil metálico de muestras de pimentón pertenecientes a las denominaciones de origen Pimentón de la Vera, localizada en la Comunidad Autónoma de Extremadura, y Pimentón de Murcia, determinando los contenidos de aluminio, boro, calcio, cobre, estroncio, fósforo, hierro, magnesio, manganeso, níquel, plomo, potasio, sodio y zinc.

2. Se ha realizado un estudio comparativo, utilizando ANOVA y la prueba de la genuina diferencia de Tukey, sobre diversos procedimientos de mineralización de la muestra, considerándose métodos por vía húmeda con distintas mezclas de ácidos. Según los resultados obtenidos y teniendo en consideración criterios de adecuación práctica, se ha concluido que el método de mineralización más adecuado es el que emplea una mezcla $\text{HNO}_3/\text{HClO}_4$.

3. Se ha puesto a punto un método de determinación de aluminio, boro, calcio, cobre, estroncio, fósforo, hierro, magnesio, manganeso, níquel, plomo, potasio, sodio y zinc en muestras de pimentón por espectrometría de emisión atómica. La veracidad se validó mediante ensayos de recuperación, obteniéndose valores estadísticamente iguales al 100% para todos los elementos considerados. La precisión se ha validado considerando la repetitividad y la precisión intermedia. Los valores de repetitividad, dados como desviación estándar relativa, varían entre 0,72% y 10,21%, y valores de precisión intermedia entre 1,17% y 11,68%. Estos valores se ajustan al criterio de aceptación de Horwitz. La linealidad que se obtiene en todos los casos está comprendida en el intervalo 97-99%. Los límites de detección se encuentran en el rango de 0,001-0,064 mg L^{-1} , mientras que los límites de cuantificación presentan valores dentro del rango 0,003-0,21 mg L^{-1} .

4. Se han determinado los contenidos de aluminio, boro, calcio, cobre, estroncio, fósforo, hierro, magnesio, manganeso, níquel, plomo, potasio, sodio y zinc en muestras de los tres tipos de pimentón, dulce, picante y agridulce, pertenecientes a las denominaciones de origen Pimentón de la Vera y Pimentón de Murcia. Los elementos potasio, fósforo, magnesio, calcio y sodio son los que

presentan concentraciones más elevadas, con valores del orden de 14500, 3500, 2000, 1900 y 400 mg kg⁻¹, respectivamente. Hierro y aluminio tienen concentraciones alrededor de 190 120 mg kg⁻¹, respectivamente. Boro, cobre, manganeso, estroncio y zinc tienen concentraciones entre 34 y 9 mg kg⁻¹. Los elementos níquel y plomo tienen concentraciones inferiores a 1 mg kg⁻¹.

5. Se ha hecho un estudio para calcular los parámetros ASTA y las coordenadas de color L*, a*, b*, c* de muestras de pimentón pertenecientes a las denominaciones de origen Pimentón de la Vera y Pimentón de Murcia. El cálculo se ha realizado a partir de los espectros de absorción UV-Vis, registrados en el rango de 250 – 800 nm. Los resultados obtenidos indican que los valores de ASTA son mayores para los pimentones picantes de Extremadura y agridulces de Murcia. En el caso de las coordenadas L*, b*, C* y H los valores mayores se corresponden con los pimentones agridulces de los dos orígenes considerados.

6. Se ha puesto a punto un método para la determinación de los capsaicinoides, capsaicina y dihidrocapsaicina en muestras de pimentón mediante HPLC, utilizando una columna de C18 y una fase móvil acetonitrilo/agua, con detección ultravioleta a 280 nm. Las condiciones cromatográficas se han optimizado mediante un diseño experimental factorial con dos niveles y tres factores.

7. El método cromatográfico se ha validado considerando la exactitud y la precisión. La exactitud se validó mediante ensayos de recuperación. Las recuperaciones obtenidas fueron del 96 y 97% para la capsaicina y dihidrocapsaicina, respectivamente. Se ha estudiado la repetitividad y la precisión intermedia, obteniéndose valores de repetitividad iguales a 3,81% y 3,59% y de precisión intermedia iguales a 4,44% y 4,27% para capsaicina y dihidrocapsaicina, respectivamente. Estos valores están dentro de los límites propuestos por Horwitz y la AOAC. Los límites de detección y cuantificación para la capsasicina son 0,06 mg L⁻¹ y 0,2 mg L⁻¹, respectivamente y para la dihidrocapsaicina 0,12 mg L⁻¹ y 0,4 mg L⁻¹. La linealidad es superior al 98% en ambos casos.

8. Se han optimizado las condiciones de extracción de capsaicina y dihidrocapsaicina en muestras de pimentón. Se utilizó acetonitrilo como el disolvente extractante, y se ha estudiado la influencia del volumen de extractante, y el tiempo y temperatura de extracción. Se utilizan cartuchos SPE de C18 para purificar los extractos obtenidos.

9. Se han determinado los contenidos de capsaicina y dihidrocapsaicina en muestras de pimentón de los tres tipos, picante, agridulce y dulce, pertenecientes a las denominaciones de origen Pimentón de la Vera y Pimentón de Murcia. El rango de concentraciones de capsaicina para las muestras de pimentón picante es de 26,4 a 919,8 mg kg⁻¹ y de 17,2 a 476,4 mg kg⁻¹ para la dihidrocapsaicina. Para las muestras agridulces el intervalo de capsaicina es 1,3 a 20,1 mg kg⁻¹ y 1,6 a 20,6 mg kg⁻¹ para la dihidrocapsaicina. Las muestras de pimentón dulce presentaron valores de capsaicina comprendidos entre 2,6 y 8,9 mg kg⁻¹, y entre 1,1 y 4,7 mg kg⁻¹ para la dihidrocapsaicina.

10. Se han calculado modelos que han permitido predecir las unidades SHU de las muestras a partir de los resultados de la concentración de capsaicinoides de los pimentones picantes, y los datos de absorbancia de obtenidos a partir de espectros UV-Vis, medidos en el rango de 215-315 nm. Con este fin, se han aplicado técnicas de regresión lineal múltiple, regresión en componentes principales, regresión por mínimos cuadrados parciales y redes neuronales. Los modelos obtenidos predicen valores SHU con una eficacia del 80%.

11. Se han aplicado técnicas de reconocimiento de patrones para la diferenciación geográfica de muestras pertenecientes a las dos denominaciones de origen de pimentón español, Pimentón de la Vera (Extremadura) y Pimentón de Murcia. Dentro de cada denominación de origen se han considerado los tipos de pimentón picante, dulce y agridulce. El estudio se ha llevado a cabo considerando distintos tipos de variables: perfil metálico, absorbancias en la zona UV-Vis, parámetro ASTA y las coordenadas CIELab, L*, a*, b*, c*.

12. Las técnicas de reconocimiento de patrones utilizadas han sido análisis en componentes principales (PCA), análisis discriminante lineal (LDA), máquinas de vectores soporte (SVM), modelado suave por analogía de clases (SIMCA) y redes neuronales artificiales (ANN). PCA se ha usado como técnica de previsualización y LDA, SVM, SIMCA y ANN para obtener modelos de clasificación.

13. Cuando se ha utilizado el perfil metálico, los elementos con mayor poder discriminante para las muestras de pimentón picante fueron cobre, fósforo, hierro, manganeso, sodio y estroncio. En el caso de las muestras de pimentón picante han sido aluminio, cobre, hierro, sodio, plomo y estroncio, y para las muestras de pimentón agridulce los elementos con mayor poder discriminante resultaron ser aluminio, calcio, hierro, sodio, plomo, estroncio, níquel, manganeso y magnesio. Los modelos más eficaces de clasificación se han obtenido con ANN, dando valores de los parámetros SENS y SPEC del 100%.

14. Los modelos de clasificación obtenidos a partir de los parámetros ASTA, L*, a* y b* han sido evaluados en función de los parámetros SENS y SPEC. Los valores, en muchos casos, no superan el 70%. Cuando se han utilizado redes neuronales perceptrón multicapa (MLP-ANN) se han obtenido valores de SENS y SPEC próximos al 95%.

15. Se han utilizado las absorbancias UV-Vis, en el intervalo 380 – 780 nm, como parámetros para construir modelos de clasificación para diferenciar el origen geográfico de muestras de pimentón. Los valores de SENS y SPEC obtenidos para el modelo MLP-ANN son superiores al 95%. El resto de los modelos da valores inferiores.

ANEXO 1

A1.1. Concentración (mg kg^{-1}) de Al, B, Ca, Cu, Mg, Mn, Na, Ni, P, Pb, Sr y Zn en muestras de pimentón.

A1.2. Valores de ASTA, L^* , a^* , b^* , C^* y H en muestras de pimentón.

A1.3. Concentración de C, DHC para pimentón picante en mg kg^{-1} . SHU: Unidades Scoville.

A1.4. Concentración de C, DHC para pimentón agridulce en mg kg^{-1} . SHU: Unidades Scoville.

A1.5. Concentración de C, DHC para pimentón dulce en mg kg^{-1} . SHU: Unidades Scoville.

ANEXO 2

A2.1. Distribución del estadístico Lilliefors.

A2.2. Valores de la distribución t de Student

A2.3. Valores críticos F de Fisher para una cola. v_1 : numerador; v_2 : denominador

A2.4. Valores críticos F de Fisher para dos colas. v_1 : numerador; v_2 : denominador

A2.5. Valores críticos para la prueba de Tukey.

A2.6. Distribución estándar normal. Valores correspondientes al área izquierda de z.

A2.7. Valores críticos $\chi^2_{(\alpha, \nu)}$ de la distribución χ^2_ν Chi-cuadrada.

A2.8. Valores críticos de Grubbs

A2.9. Valores del estadístico C de Cochran.

ANEXO 1

A1.1. Concentración (mg kg⁻¹) de Al, B, Ca, Cu, Mg, Mn, Na, Ni, P, Pb, Sr y Zn en muestras de pimentón

Muestra	Al	B	Ca	Cu	Fe	K	Mg	Mn	Na	Ni	P	Pb	Sr	Zn
EX01 ^a	133,9(1)	10,9(1)	1445,8(8)	8,62(7)	208,6(5)	15475,8(9)	2288,1(3)	21,5(2)	542,9(6)	1,06(5)	3954,8(9)	0,605(6)	6,89(1)	25,0(1)
EX02 ^a	129,8(9)	11,4(3)	1566,2(1)	9,47(2)	211,1(8)	16038,1(2)	2269,7(6)	23,4(3)	559,9(4)	1,12(3)	3961,3(5)	0,656(5)	7,65(1)	26,3(7)
EX03 ^a	144,9(9)	17,5(5)	2085,1(2)	19,8(7)	155,8(6)	15272,7(4)	2129,3(6)	61,8(7)	426,8(7)	3,29(8)	3473,7(9)	8,13(5)	16,32(5)	60,5(5)
EX04 ^a	157,4(4)	21,8(4)	2183,9(3)	19,4(2)	166,4(8)	11673,7(3)	2198,6(5)	66,8(4)	435,1(6)	3,40(9)	3555,8(3)	8,09(9)	16,18(2)	59,0(5)
EX05 ^a	120,6(6)	23,8(6)	1796,6(2)	19,4(6)	171,6(4)	12501,6(2)	2319,2(4)	36,6(8)	381,0(1)	0,688(6)	4139,9(8)	0,280(4)	16,10(9)	56,6(7)
EX06 ^a	114,2(5)	23,3(4)	1679,2(6)	19,6(2)	165,5(8)	10318,2(4)	2237,7(1)	37,3(5)	361,7(8)	0,673(5)	4007,4(1)	0,134(3)	15,87(7)	52,5(1)
EX07 ^a	118,8(2)	15,3(3)	1795,7(8)	15,1(5)	185,4(5)	14998,2(6)	2072,5(1)	52,1(3)	404,6(4)	0,755(9)	3460,8(8)	8,44(5)	12,82(8)	45,0(5)
EX08 ^a	124,4(9)	15,2(9)	1768,0(8)	17,1(8)	191,0(9)	15335,8(2)	1964,4(3)	54,5(1)	404,9(9)	0,696(8)	3263,0(9)	9,27(3)	14,40(2)	50,7(3)
EX09 ^a	139,0(1)	13,7(6)	2232,2(8)	7,87(8)	229,3(6)	20213,2(6)	2108,4(5)	23,2(6)	434,9(1)	1,23(7)	3353,7(7)	0,917(1)	12,92(6)	22,8(9)
EX10 ^a	126,2(3)	10,1(9)	2169,2(4)	7,44(5)	210,9(1)	18633,2(1)	2107,1(9)	22,3(6)	399,5(2)	1,09(2)	3285,0(9)	0,321(5)	11,24(7)	21,1(6)
EX11 ^a	91,1(5)	8,98(7)	1594,9(5)	9,82(7)	160,2(1)	14937,6(2)	2219,0(4)	27,5(7)	301,9(9)	2,18(8)	3612,2(9)	0,439(9)	7,44(8)	21,3(3)
EX12 ^a	91,5(2)	8,4(3)	1590,7(6)	9,86(1)	159,5(3)	20084,5(3)	2096,6(5)	24,4(5)	306,9(9)	2,16(9)	3424,3(9)	0,466(4)	7,08(2)	18,3(6)
EX13 ^a	110,8(4)	6,68(5)	1905,6(4)	9,07(6)	153,6(2)	15093,8(5)	2118,8(5)	23,0(2)	373,7(1)	0,843(9)	3932,8(4)	0,103(4)	9,05(6)	33,9(8)
EX14 ^a	130,7(4)	6,95(5)	2017,2(4)	8,88(9)	160,7(3)	12994,6(8)	2200,8(1)	22,4(5)	362,2(1)	0,868(8)	3896,7(4)	0,094(5)	9,25(1)	33,9(8)
EX15 ^a	115,5(5)	6,36(7)	1722,0(2)	8,59(2)	150,4(1)	15428,7(2)	1988,2(3)	21,4(4)	349,4(4)	0,804(8)	3601,0(5)	0,260(2)	8,37(8)	36,1(4)
EX16 ^a	135,0(8)	6,83(7)	1918,2(2)	8,96(8)	164,5(5)	17452,0(9)	2207,1(2)	22,2(2)	396,2(1)	0,984(1)	3805,0(5)	0,165(2)	9,13(1)	34,3(7)
EX17 ^a	104,2(3)	6,27(8)	1833,8(1)	9,81(8)	183,7(1)	11462,5(5)	2227,5(1)	32,3(4)	225,2(3)	1,66(2)	3947,6(2)	0,044(5)	6,62(2)	30,4(5)
EX18 ^a	106,6(6)	5,77(4)	1890,3(7)	9,93(2)	187,6(7)	12794,0(1)	2163,6(8)	32,0(4)	228,9(1)	1,55(7)	3809,7(2)	0,325(7)	6,76(4)	33,0(3)
EX19 ^a	227,3(1)	6,03(1)	1933,0(5)	6,53(4)	241,9(7)	15138,9(6)	2345,6(9)	19,4(9)	437,3(8)	0,492(5)	3715,6(7)	0,409(2)	8,79(4)	27,4(4)
EX20 ^a	210,9(1)	7,65(7)	2066,1(3)	6,49(2)	246,9(4)	12390,4(4)	2288,3(2)	19,1(9)	409,3(5)	0,487(3)	3626,5(6)	0,131(8)	8,20(1)	27,5(5)
EX21 ^a	178,4(6)	15,1(7)	2421,7(9)	17,5(9)	214,6(6)	14189,6(3)	2204,7(2)	68,4(8)	205,3(1)	2,19(3)	3459,4(7)	19,02(3)	11,45(8)	47,7(4)
EX22 ^a	174,5(7)	14,8(3)	2349,0(9)	20,6(3)	225,2(7)	15460,3(6)	2189,9(2)	68,5(3)	201,0(3)	2,09(6)	3364,5(5)	21,28(1)	11,80(3)	56,6(5)
EX23 ^a	132,7(1)	11,2(9)	1281,7(3)	11,0(5)	182,6(1)	16208,7(5)	2062,7(9)	31,3(5)	300,8(5)	0,774(2)	3810,8(3)	0,478(8)	9,22(7)	27,8(3)
EX24 ^a	117,8(2)	10,1(5)	1112,6(1)	10,2(1)	167,5(9)	16810,4(3)	1950,6(6)	29,9(2)	279,0(5)	0,875(9)	3555,2(9)	0,266(6)	8,45(4)	30,3(9)
EX25 ^d	188,9(3)	14,7(9)	1764,8(9)	10,9(3)	264,9(1)	16223,8(8)	2128,4(9)	28,6(7)	531,1(6)	2,94(2)	3495,9(1)	3,32(3)	7,48(9)	28,1(3)
EX26 ^d	195,4(8)	14,6(2)	1889,7(1)	9,88(4)	257,9(6)	16003,8(8)	2124,7(6)	28,1(5)	530,8(3)	2,83(8)	3508,3(9)	4,15(7)	6,48(1)	24,5(6)

A1.1. Continuación.

Muestra	Al	B	Ca	Cu	Fe	K	Mg	Mn	Na	Ni	P	Pb	Sr	Zn
EX27 ^d	168,3(9)	12,5(6)	1732,9(4)	11,6(5)	294,9(7)	13576,0(1)	2283,3(5)	19,7(4)	344,3(3)	1,03(3)	4128,5(4)	0,196(6)	23,35(2)	29,4(2)
EX28 ^d	171,6(4)	12,8(7)	1456,5(7)	11,6(9)	295,4(1)	13620,9(2)	2314,7(2)	19,7(8)	350,6(7)	1,07(4)	4188,9(1)	0,234(9)	25,78(1)	31,8(7)
EX29 ^d	244,3(2)	27,9(6)	1877,5(5)	20,0(9)	229,3(7)	15872,4(8)	1976,1(1)	55,5(8)	482,5(3)	2,51(7)	3544,6(5)	3,50(6)	15,37(4)	68,9(8)
EX30 ^d	245,7(2)	22,9(7)	2329,4(4)	20,0(3)	222,6(3)	15486,6(7)	2215,3(4)	55,9(7)	483,7(5)	2,46(2)	3569,8(5)	3,05(1)	15,35(3)	69,6(6)
EX31 ^d	172,6(9)	21,7(7)	2465,1(5)	15,1(1)	230,8(6)	11859,9(5)	2460,9(4)	97,6(3)	410,8(8)	0,626(7)	3653,5(2)	0,096(5)	29,09(9)	62,7(4)
EX32 ^d	139,3(9)	22,0(6)	2185,0(6)	14,1(7)	209,2(2)	11174,6(3)	2292,8(1)	88,1(2)	384,5(1)	0,570(7)	3428,3(6)	0,272(6)	26,06(9)	56,7(3)
EX33 ^d	119,7(1)	15,7(8)	2103,7(7)	19,1(1)	202,0(3)	15064,7(7)	2154,8(4)	47,5(1)	393,6(9)	2,01(3)	3185,0(1)	15,41(9)	19,67(3)	48,4(5)
EX34 ^d	95,5(3)	18,0(4)	2389,4(1)	16,4(8)	137,7(2)	14704,7(2)	2210,6(8)	39,5(7)	367,8(2)	1,77(6)	3191,9(8)	18,98(6)	15,07(2)	40,9(3)
EX35 ^d	105,4(6)	9,11(1)	1900,0(1)	5,98(5)	158,8(7)	16256,8(9)	2058,0(9)	22,0(7)	294,0(8)	0,615(9)	3636,5(1)	0,129(5)	7,25(4)	29,7(6)
EX36 ^d	99,1(1)	7,29(8)	2194,0(6)	6,95(5)	157,2(3)	14917,8(5)	2172,9(7)	23,2(6)	292,6(7)	0,611(9)	3811,6(2)	0,048(7)	7,82(8)	32,7(2)
EX37 ^d	79,6(2)	17,8(7)	1649,4(8)	9,64(5)	177,2(2)	14691,6(2)	2276,6(1)	22,8(4)	366,0(4)	1,14(8)	3721,0(3)	0,434(2)	8,88(6)	18,8(2)
EX38 ^d	82,7(3)	8,92(9)	1545,9(9)	9,51(5)	176,8(5)	14580,2(7)	2277,0(8)	22,4(5)	338,9(7)	1,23(6)	3768,3(8)	0,509(6)	8,36(1)	19,7(8)
EX39 ^d	125,6(7)	15,4(4)	2221,4(9)	12,2(5)	176,4(8)	14649,0(3)	2173,4(2)	25,7(7)	299,3(7)	0,966(5)	3968,2(7)	0,308(5)	8,55(5)	34,1(7)
EX40 ^d	138,6(6)	14,9(6)	1988,4(5)	11,5(9)	178,9(8)	13496,1(7)	2237,7(8)	25,3(8)	288,7(5)	1,08(7)	3920,7(4)	0,231(3)	8,86(4)	33,1(1)
EX41 ^d	106,9(9)	20,3(1)	2215,9(2)	19,2(1)	155,5(6)	17571,3(9)	2203,8(8)	44,8(6)	269,2(1)	0,947(6)	3638,0(8)	8,37(1)	15,90(4)	62,1(2)
EX42 ^d	146,8(3)	17,9(2)	1880,9(4)	19,8(9)	175,9(9)	16775,8(8)	2135,2(1)	44,5(2)	270,4(3)	1,26(7)	3541,4(4)	7,98(1)	15,65(9)	63,7(8)
EX43 ^d	143,4(6)	7,44(1)	1962,5(7)	9,10(2)	182,8(7)	13692,7(3)	2273,1(2)	28,5(3)	252,0(6)	0,452(8)	4266,0(4)	0,039(6)	7,96(3)	38,1(5)
EX44 ^d	145,5(4)	7,85(9)	2136,4(8)	9,24(1)	183,1(1)	13144,4(4)	2279,6(8)	27,9(8)	237,9(4)	0,452(5)	4158,5(3)	0,360(8)	7,89(7)	37,3(8)
EX45 ^d	100,1(8)	15,1(6)	1376,8(6)	14,4(3)	109,7(1)	15225,1(3)	1933,9(8)	26,0(5)	195,7(8)	0,213(7)	3228,8(5)	6,36(3)	7,07(7)	57,2(6)
EX46 ^d	98,8(6)	7,66(7)	1138,4(1)	10,5(9)	83,8(8)	14763,3(8)	1483,4(1)	20,7(8)	154,6(6)	0,168(1)	2498,7(8)	6,36(3)	6,37(5)	48,0(1)
EX47 ^d	102,4(3)	9,91(7)	1068,1(8)	8,40(7)	154,9(5)	17581,2(4)	1879,1(9)	20,1(3)	316,3(8)	0,497(6)	3590,0(4)	0,093(5)	6,29(2)	32,6(4)
EX48 ^d	100,8(8)	11,5(4)	1062,6(9)	8,67(8)	156,4(4)	14636,9(8)	1842,0(8)	20,4(3)	308,7(3)	0,413(5)	3560,0(6)	0,033(5)	5,69(8)	29,7(3)
EX49 ^p	124,4(2)	8,88(6)	2120,7(8)	9,16(5)	168,3(7)	20482,4(9)	2299,5(8)	20,3(9)	343,4(6)	0,711(3)	3812,0(7)	6,88(2)	6,15(4)	24,8(9)
EX50 ^p	126,8(2)	10,5(4)	1832,5(8)	7,89(8)	153,5(6)	18911,5(8)	1952,8(2)	17,7(8)	278,9(1)	0,521(5)	3321,3(2)	6,13(8)	5,48(9)	18,9(3)
EX51 ^p	127,6(3)	10,4(4)	858,8(4)	9,18(1)	235,7(4)	13131,0(6)	2176,5(8)	15,3(6)	479,5(8)	1,05(7)	3581,1(9)	0,408(8)	7,26(1)	21,1(6)
EX52 ^p	125,8(5)	9,11(2)	832,9(7)	9,32(3)	235,7(8)	13189,5(1)	2181,9(8)	15,3(4)	478,8(6)	0,949(9)	3660,7(1)	0,358(9)	7,87(9)	19,0(1)

A1.1. Continuación.

Muestra	Al	B	Ca	Cu	Fe	K	Mg	Mn	Na	Ni	P	Pb	Sr	Zn
EX53 ^p	128,0(6)	21,2(3)	2240,9(9)	15,7(7)	123,1(4)	13471,2(4)	2191,5(6)	31,9(6)	356,9(4)	0,833(7)	3581,5(2)	0,347(1)	9,21(6)	45,5(2)
EX54 ^p	147,8(4)	16,4(1)	2033,0(1)	16,4(6)	133,0(5)	14177,3(2)	1997,0(9)	31,9(3)	348,3(1)	0,769(3)	3552,9(4)	0,154(9)	9,45(4)	49,9(2)
EX55 ^p	104,6(7)	23,0(6)	2151,2(1)	17,5(3)	213,3(6)	10815,8(1)	2249,8(2)	42,9(4)	288,5(3)	0,818(1)	3769,2(1)	0,675(3)	16,24(4)	57,1(8)
EX56 ^p	98,3(6)	23,5(2)	2170,6(2)	16,9(1)	198,7(4)	10639,6(8)	2239,5(6)	41,9(9)	291,6(8)	0,737(5)	3804,4(8)	1,33(7)	15,54(9)	62,2(3)
EX57 ^p	116,2(1)	18,0(3)	2277,1(8)	20,6(4)	191,4(9)	13991,4(6)	2389,9(6)	55,6(4)	304,8(6)	2,03(8)	3555,0(9)	9,58(5)	21,28(3)	51,2(5)
EX58 ^p	123,2(2)	16,8(5)	2318,2(5)	20,6(9)	210,5(7)	12664,7(4)	2292,3(1)	54,4(1)	287,0(9)	2,13(9)	3432,2(7)	9,93(1)	19,85(6)	48,9(5)
EX59 ^p	83,5(6)	13,1(6)	2125,7(2)	7,71(8)	148,3(8)	19703,1(9)	2191,6(5)	23,2(5)	371,2(4)	0,735(9)	3667,5(5)	0,461(2)	9,36(3)	22,3(7)
EX60 ^p	95,9(7)	13,5(8)	2126,8(4)	7,36(7)	153,3(6)	14973,6(5)	2172,4(2)	22,6(7)	341,2(4)	0,814(9)	3606,5(6)	0,348(9)	9,51(7)	21,5(1)
EX61 ^p	97,1(2)	13,4(3)	1490,7(5)	18,3(7)	201,1(2)	14345,4(5)	2152,6(9)	27,8(1)	432,9(1)	1,36(4)	3384,9(1)	7,98(1)	15,72(1)	44,8(3)
EX62 ^p	104,2(3)	14,4(9)	1685,5(2)	18,3(2)	207,7(3)	14485,8(5)	2143,7(1)	27,8(8)	427,3(5)	1,25(1)	3412,1(7)	8,89(3)	14,85(7)	35,3(1)
EX63 ^p	160,1(5)	13,4(3)	2054,0(2)	16,4(1)	158,3(2)	15591,6(5)	1976,3(8)	43,5(1)	275,3(4)	1,68(4)	3169,5(6)	9,27(5)	13,90(2)	53,6(5)
EX64 ^p	136,9(3)	16,1(6)	2286,1(4)	15,8(8)	147,1(6)	15163,8(4)	1992,1(4)	45,2(5)	279,8(8)	1,59(8)	3208,9(8)	11,05(9)	14,05(5)	54,9(1)
EX65 ^p	128,0(6)	12,8(7)	2157,2(1)	9,11(5)	158,6(6)	13047,0(7)	2071,7(7)	24,5(3)	302,7(2)	1,43(5)	3592,0(1)	0,921(3)	7,72(7)	28,5(5)
EX66 ^p	116,6(2)	14,4(8)	2436,8(3)	9,22(2)	164,3(2)	15353,2(7)	2201,6(6)	24,3(2)	325,3(7)	1,65(7)	3950,1(4)	0,282(4)	7,65(1)	28,7(6)
EX67 ^p	113,7(4)	18,8(8)	2028,4(6)	17,5(4)	209,3(2)	14959,1(4)	2169,6(5)	49,7(9)	496,6(4)	1,61(4)	3340,5(7)	13,12(6)	17,31(2)	53,9(3)
EX68 ^p	117,8(6)	20,2(4)	2033,6(9)	17,0(8)	191,5(7)	13192,3(6)	2134,8(2)	49,2(3)	474,9(2)	1,65(4)	3349,3(1)	16,81(8)	17,02(6)	49,2(3)
EX69 ^p	104,3(7)	9,0(3)	1003,5(4)	8,87(8)	167,4(1)	13886,2(5)	1866,5(5)	22,6(1)	310,9(4)	0,943(4)	3286,0(4)	0,901(9)	7,97(9)	24,5(1)
EX70 ^p	118,8(6)	9,87(6)	1190,8(4)	9,66(5)	184,4(2)	15998,8(6)	2051,5(2)	24,2(1)	370,3(8)	2,13(7)	3520,5(1)	2,85(4)	8,63(2)	24,0(7)
EX71 ^p	85,3(4)	7,01(3)	1900,2(8)	7,76(4)	134,7(2)	23416,4(3)	2212,3(6)	20,0(9)	295,4(2)	0,625(3)	3800,5(1)	0,427(2)	6,75(5)	26,1(4)
EX72 ^p	84,6(5)	13,3(5)	1839,7(5)	7,97(3)	130,0(7)	13510,4(7)	2330,0(6)	20,5(6)	255,4(5)	0,672(3)	3892,5(8)	0,095(8)	7,30(7)	27,4(5)
MU01 ^a	44,5(3)	5,61(9)	1214,4(8)	10,2(2)	107,5(9)	14654,5(2)	1737,8(3)	9,50(9)	541,8(9)	0,587(4)	3643,8(9)	0,140(7)	11,30(8)	32,6(1)
MU02 ^a	50,7(6)	5,46(9)	1170,9(9)	10,6(6)	106,3(3)	16529,8(8)	1744,2(4)	9,05(8)	557,5(1)	0,549(9)	3669,9(2)	0,194(6)	11,41(1)	29,7(7)
MU03 ^a	41,1(9)	6,29(6)	1293,8(3)	10,3(7)	99,8(8)	14387,3(8)	1774,5(1)	9,54(3)	545,6(2)	0,576(9)	3725,4(2)	0,010(9)	11,50(8)	34,6(3)
MU04 ^a	44,3(6)	6,07(2)	1208,3(7)	11,3(4)	105,1(6)	15116,3(4)	1833,1(9)	10,0(8)	570,4(4)	0,570(7)	3721,8(7)	0,144(4)	11,78(1)	30,9(7)
MU05 ^a	42,1(8)	5,58(1)	1121,7(3)	10,4(7)	103,2(2)	14525,9(4)	1683,2(2)	9,34(5)	522,7(1)	0,543(9)	3590,5(8)	0,011(3)	10,78(6)	26,8(9)
MU06 ^a	52,3(4)	5,97(3)	1180,5(7)	10,8(1)	115,9(1)	17547,5(9)	1785,4(6)	9,48(5)	573,2(2)	0,667(8)	3771,4(9)	0,035(9)	11,45(9)	34,0(5)

A1.1. Continuación.

Muestra	Al	B	Ca	Cu	Fe	K	Mg	Mn	Na	Ni	P	Pb	Sr	Zn
MU07 ^a	114,8(4)	22,9(6)	2174,1(9)	22,6(7)	223,2(4)	14981,1(6)	1937,6(6)	26,5(3)	924,7(9)	0,756(6)	3402,1(3)	5,45(9)	30,97(6)	59,0(3)
MU08 ^a	136,8(2)	14,2(1)	2117,8(3)	20,7(9)	264,4(1)	13150,3(2)	1911,8(9)	26,5(5)	947,6(9)	0,868(8)	3418,0(4)	7,72(7)	30,02(1)	69,0(4)
MU09 ^a	131,6(4)	23,6(1)	2132,3(1)	22,0(9)	261,1(3)	13606,2(2)	1986,4(4)	27,7(2)	940,9(6)	0,877(2)	3530,4(6)	9,09(7)	32,13(1)	67,4(3)
MU10 ^a	33,5(7)	12,4(4)	1099,8(9)	11,1(1)	108,8(3)	14716,0(3)	1807,7(4)	10,9(6)	598,0(6)	0,589(5)	3620,2(9)	0,041(1)	10,77(9)	29,4(5)
MU11 ^a	26,4(8)	11,9(7)	1036,3(2)	11,1(3)	99,3(8)	13913,0(5)	1767,4(3)	10,6(7)	538,0(8)	0,550(7)	3529,7(3)	0,099(4)	10,64(7)	26,8(9)
MU12 ^a	32,4(1)	12,9(1)	1030,6(2)	9,21(2)	106,6(6)	10263,6(1)	1659,0(4)	10,3(6)	589,6(7)	0,544(4)	3274,7(2)	0,023(4)	10,60(3)	30,4(4)
MU13 ^a	46,2(4)	13,5(3)	1241,9(3)	11,8(2)	105,9(7)	12471,1(5)	1780,6(9)	10,2(7)	565,2(3)	0,625(4)	3817,8(1)	0,454(5)	12,15(5)	30,5(2)
MU14 ^a	44,1(8)	13,4(7)	1181(1)	11,9(1)	110,0(9)	12270,5(7)	1775,6(2)	10,7(9)	576,5(3)	0,603(3)	3665,1(4)	0,110(7)	9,04(1)	23,9(5)
MU15 ^a	40,1(5)	13,3(2)	1271,4(8)	11,7(4)	103,0(5)	7691,0(2)	1802,0(4)	10,7(7)	494,4(8)	0,681(7)	3812,1(1)	0,038(9)	12,83(2)	33,1(6)
MU16 ^a	28,0(1)	6,81(3)	1720,3(5)	9,95(7)	87,5(7)	14285,5(4)	1772,9(5)	9,47(4)	286,7(6)	0,612(9)	3550,2(5)	0,006(5)	17,85(1)	30,5(7)
MU17 ^a	34,6(9)	6,53(2)	1739,8(7)	10,1(3)	99,1(4)	14569,6(2)	1796,0(2)	9,13(5)	298,1(4)	0,631(6)	3616,3(4)	0,226(1)	18,25(2)	29,5(8)
MU18 ^a	34,1(4)	7,92(9)	1913,1(1)	11,2(7)	98,4(3)	15820,3(6)	1985,8(6)	10,7(9)	331,5(4)	0,711(1)	3864,4(1)	0,011(5)	20,50(5)	29,7(6)
MU19 ^a	29,7(9)	6,86(4)	1519,9(3)	9,44(9)	85,3(8)	13330,6(8)	1642,3(2)	9,23(1)	277,8(5)	0,565(7)	3154,4(1)	0,021(6)	17,01(2)	27,1(7)
MU20 ^a	31,9(7)	7,19(9)	1870,1(2)	10,3(7)	96,5(6)	14764,0(2)	1842,1(1)	9,61(2)	319,8(1)	0,622(2)	3648,4(4)	0,000(8)	18,86(1)	27,7(6)
MU21 ^a	43,7(1)	8,16(2)	2130,8(6)	12,3(7)	115,5(2)	13182,1(9)	2190,3(7)	10,9(6)	342,8(9)	0,753(4)	4226,3(3)	0,232(7)	22,25(4)	32,7(7)
MU22 ^d	117,0(8)	12,8(4)	1688,5(5)	10,5(5)	231,8(8)	17349,4(3)	1874,3(5)	14,3(6)	683,1(2)	0,987(6)	3404,2(9)	4,92(3)	16,52(3)	22,3(3)
MU23 ^d	115,2(2)	11,5(5)	1605,4(8)	10,8(2)	227,2(3)	18129,4(3)	1889,7(5)	14,4(1)	684,2(6)	1,12(7)	3406,0(9)	0,373(9)	16,25(9)	23,9(2)
MU24 ^d	103,7(9)	11,2(3)	1498(9)	9,37(6)	228,7(4)	16710,8(2)	1757,6(3)	12,5(9)	634,2(4)	0,930(4)	3154,1(3)	0,102(4)	14,23(4)	14,9(4)
MU25 ^d	98,1(6)	11,4(9)	1673,2(1)	9,33(7)	182,5(2)	18595,6(5)	1968,3(9)	12,5(2)	606,0(8)	0,843(8)	3484,2(5)	0,350(8)	10,79(7)	24,9(2)
MU26 ^d	94,8(8)	15,0(3)	1710,8(6)	9,25(4)	188,5(2)	18730,0(4)	2026,6(5)	10,7(1)	613,4(8)	0,819(8)	3575,2(8)	0,318(7)	9,59(2)	18,3(6)
MU27 ^d	105,0(9)	14,2(1)	1811,7(5)	9,41(7)	197,2(1)	20157,8(8)	2011,8(1)	12,7(9)	638,4(7)	0,815(5)	3559,9(9)	1,93(1)	10,07(6)	24,9(9)
MU28 ^d	130,3(3)	20,5(5)	2283,3(3)	20,3(8)	208,0(5)	9169,1(5)	1942,8(4)	25,4(1)	438,8(3)	0,808(9)	3458,5(3)	0,115(3)	33,74(1)	46,8(7)
MU29 ^d	139,6(1)	7,38(3)	2272,5(2)	21,3(2)	214,9(3)	15435,2(2)	1942,1(7)	25,7(4)	496,0(5)	0,820(4)	3329,0(4)	0,030(7)	34,45(1)	48,6(7)
MU30 ^d	120,9(7)	20,8(3)	2384(2)	20,7(2)	210,3(5)	9722,2(5)	1931,5(7)	24,9(1)	446,2(5)	0,870(9)	3400,1(2)	0,174(5)	34,86(8)	48,5(4)
MU31 ^d	132,8(8)	18,3(9)	2150,3(4)	22,2(6)	217,1(2)	16145,7(4)	1894,3(6)	25,6(7)	509,1(4)	1,34(4)	3435,7(9)	1,99(4)	35,36(3)	53,1(1)
MU32 ^d	141,8(1)	15,8(1)	2066,2(8)	20,7(1)	203,2(9)	14492,0(8)	1776,7(7)	24,0(5)	518,5(7)	1,00(1)	3191,2(9)	0,516(4)	32,26(3)	62,2(2)

A1.1. Continuación.

Muestra	Al	B	Ca	Cu	Fe	K	Mg	Mn	Na	Ni	P	Pb	Sr	Zn
MU33 ^d	119,5(7)	19,8(3)	2182,2(2)	20,3(2)	211,3(2)	13624,6(8)	1922,4(7)	25,2(1)	493,5(4)	0,960(6)	3478,5(4)	1,97(2)	33,28(4)	53,4(8)
MU34 ^d	167,1(5)	10,8(2)	1987,5(6)	10,0(6)	375,6(6)	15194,5(3)	2487,8(3)	17,9(7)	432,2(1)	1,000(7)	3093,9(8)	0,490(5)	19,09(3)	15,1(7)
MU35 ^d	190,0(9)	10,3(2)	1909,6(9)	10,2(5)	402,9(7)	15820,8(9)	2475,1(7)	17,4(4)	419,9(5)	1,08(1)	3080,8(1)	0,429(4)	17,31(5)	14,9(5)
MU36 ^d	173,7(6)	10,7(6)	1896,2(5)	10,1(9)	393,5(4)	14394,9(7)	2489,3(7)	17,2(5)	418,9(6)	0,932(6)	3098,4(5)	0,679(5)	18,20(4)	15,5(8)
MU37 ^d	276,3(8)	15,6(8)	2275,3(9)	21,1(7)	362,3(7)	17185,0(9)	2622,1(6)	34,6(9)	427,5(9)	0,874(1)	3204,9(2)	0,369(1)	38,50(2)	43,7(7)
MU38 ^d	227,2(5)	15,6(4)	2340,8(5)	21,0(6)	370,6(2)	15352,5(1)	2566,7(9)	33,3(2)	423,8(9)	0,827(4)	3145,8(3)	0,462(4)	36,52(9)	41,1(8)
MU39 ^d	220,4(8)	15,9(2)	2442,0(7)	20,1(4)	374,6(3)	14833,7(3)	2656,5(8)	35,8(4)	438,4(8)	0,927(6)	3238,1(7)	0,749(5)	36,93(5)	43,4(4)
MU40 ^d	186,2(9)	15,2(8)	1946,2(9)	21,8(8)	299,4(3)	16083,9(1)	2284,0(8)	30,2(3)	880,6(8)	1,43(8)	3396,2(1)	10,71(5)	30,90(1)	55,6(8)
MU41 ^d	180,9(3)	11,1(2)	1726,2(5)	20,6(7)	260,1(9)	13250,9(5)	2061,9(3)	27,5(4)	768,6(8)	1,29(2)	3118,1(6)	9,55(2)	28,46(4)	48,6(1)
MU42 ^d	170,8(6)	14,1(9)	1798,9(7)	20,3(9)	281,3(7)	15687,0(2)	2235,7(3)	29,7(3)	820,7(9)	1,39(5)	3415,4(7)	10,24(4)	29,82(5)	43,3(4)
MU43 ^p	133,9(7)	13,8(5)	1494,3(4)	9,58(6)	189,4(4)	13650,0(8)	1984,9(9)	15,4(8)	569,4(7)	0,948(5)	3240,7(1)	0,497(6)	10,90(4)	17,6(9)
MU44 ^p	132,4(9)	13,5(1)	1279,8(2)	9,46(3)	193,8(1)	13047,7(2)	2014,4(1)	15,1(1)	479,4(7)	0,980(5)	3372,1(4)	0,694(4)	11,28(5)	18,5(1)
MU45 ^p	131,6(4)	13,2(8)	1312,3(1)	9,86(9)	194,5(6)	14738,1(2)	1967,4(2)	15,5(2)	479,8(8)	0,953(2)	3391,9(4)	0,287(6)	11,15(4)	21,6(1)
MU46 ^p	111,0(9)	9,98(2)	1442,4(2)	9,43(6)	182,9(6)	19646,5(1)	2405,9(7)	12,6(9)	413,9(1)	1,08(6)	3828,5(6)	0,520(8)	8,49(7)	14,2(1)
MU47 ^p	108,9(4)	9,08(1)	1439,6(2)	9,39(5)	178,6(9)	19765,2(6)	2363,8(6)	13,1(4)	403,1(2)	1,06(9)	3855,8(2)	0,503(3)	8,60(5)	17,0(1)
MU48 ^p	106,8(3)	9,54(2)	1420,7(1)	9,11(9)	178,4(3)	19275,8(6)	2260,1(4)	13,2(7)	407,1(1)	1,02(1)	3702,5(8)	0,395(7)	8,76(5)	15,2(7)
MU49 ^p	201,7(9)	19,3(2)	2551,5(1)	20,2(3)	309,5(6)	13107,2(7)	2237,0(4)	34,2(2)	416,5(9)	1,24(4)	3630,8(4)	0,305(8)	33,86(8)	59,0(3)
MU50 ^p	207,9(9)	22,7(3)	2505,4(4)	22,6(3)	329,1(3)	10862,3(6)	2439,5(3)	35,7(8)	426,7(6)	1,48(7)	3958,5(6)	0,444(9)	36,81(5)	58,8(3)
MU51 ^p	210,8(9)	22,2(5)	2550,4(7)	21,2(1)	294,6(1)	11784,3(9)	2460,5(7)	36,8(3)	434,0(5)	1,42(1)	4074,4(7)	0,405(8)	35,65(1)	51,9(6)
MU52 ^p	220,0(3)	17,5(9)	2536,0(9)	21,8(6)	307,4(5)	14507,0(9)	2409,3(9)	36,4(6)	452,3(2)	1,30(6)	3829,0(4)	0,268(3)	35,03(2)	51,8(9)
MU53 ^p	138,7(2)	0,261(6)	1689,7(7)	14,0(9)	169,9(3)	9226,0(7)	1703,4(4)	25,5(7)	524,6(5)	0,831(5)	2685,5(6)	0,364(6)	25,75(6)	36,5(7)
MU54 ^p	219,1(8)	16,1(3)	2343,1(1)	21,8(4)	327,3(4)	14010,9(4)	2418,3(7)	36,0(6)	447,1(5)	1,28(8)	3898,0(3)	0,537(9)	34,84(8)	53,9(4)
MU55 ^p	230,0(9)	12,1(2)	2700,2(1)	21,2(9)	343,5(7)	14129,3(6)	2269,6(6)	50,9(3)	176,7(2)	1,50(2)	3131,8(2)	5,48(3)	11,09(7)	49,8(3)
MU56 ^p	203,7(5)	11,1(1)	2614,6(3)	19,7(3)	330,1(3)	13144,4(4)	2149,2(4)	49,6(4)	168,2(1)	1,43(5)	2917,1(1)	9,16(2)	11,05(4)	51,1(1)
MU57 ^p	217,3(1)	10,8(6)	2625,5(1)	21,5(8)	343,0(9)	14361,6(7)	2289,7(2)	50,0(4)	185,4(2)	1,56(3)	3139,6(7)	7,88(9)	11,52(7)	60,8(5)

A1.1. Continuación.

Muestra	Al	B	Ca	Cu	Fe	K	Mg	Mn	Na	Ni	P	Pb	Sr	Zn
MU58 ^P	203,1(8)	15,7(2)	2578,3(4)	21,3(9)	283,0(6)	14163,5(1)	2264,7(6)	52,2(5)	202,7(8)	1,33(1)	3218,7(5)	0,263(6)	10,28(7)	51,4(1)
MU59 ^P	208,9(3)	15,8(5)	2590,3(6)	22,6(4)	294,3(8)	13887,8(7)	2330,9(5)	53,0(4)	200,7(5)	1,23(3)	3301,5(2)	0,106(7)	11,36(5)	48,1(5)
MU60 ^P	235,8(1)	13,0(1)	2789,8(4)	20,0(2)	306,6(9)	14480,0(1)	2339,0(3)	53,0(6)	206,5(6)	1,26(7)	3355,1(3)	0,098(5)	10,07(3)	45,3(2)
MU61 ^P	47,9(6)	12,2(1)	1015,0(1)	11,5(6)	123,9(3)	13587,0(5)	1885,5(3)	12,3(9)	354,9(7)	0,849(3)	3505,8(6)	0,443(9)	15,49(5)	27,8(9)
MU62 ^P	43,0(9)	11,5(7)	1011,8(7)	10,9(4)	123,3(2)	16127,2(5)	1899,5(9)	12,3(7)	377,6(3)	0,891(3)	3564,8(3)	0,091(3)	14,11(1)	26,7(7)
MU63 ^P	52,8(8)	12,2(2)	1123,9(9)	11,7(1)	133,9(5)	14568,6(7)	1933,1(4)	12,9(2)	358,9(5)	1,06(2)	3545,5(2)	0,225(6)	16,22(1)	29,6(1)
MU64 ^P	110,7(2)	9,11(4)	2181,0(1)	11,9(3)	194,7(8)	18577,1(7)	2252,8(1)	14,4(6)	701,9(8)	1,05(2)	4092,6(3)	0,203(3)	15,97(2)	30,4(2)
MU65 ^P	118,4(4)	6,95(2)	2062,1(9)	10,8(1)	198,7(3)	15009,0(6)	2078,6(2)	13,6(5)	617,2(6)	0,907(9)	3961,5(3)	0,109(7)	14,92(8)	25,8(5)
MU66 ^P	109,2(8)	7,01(9)	1957,2(8)	11,0(8)	171,0(2)	15690,0(8)	2044,9(1)	13,2(5)	598,2(7)	0,921(6)	3728,3(8)	0,282(8)	15,26(1)	24,7(5)
MU67 ^P	96,5(8)	8,45(7)	2095,9(9)	11,5(2)	152,1(2)	17950,8(5)	2119,1(8)	13,8(6)	641,4(2)	0,959(9)	3901,6(7)	0,401(6)	15,74(8)	32,5(1)
MU68 ^P	117,8(5)	7,08(6)	2201,2(3)	10,8(7)	198,2(8)	17294,4(7)	2133,8(6)	13,7(5)	647,7(2)	1,01(4)	3952,9(7)	0,154(2)	15,28(9)	26,6(1)
MU69 ^P	111,0(9)	7,73(5)	2107,0(9)	11,4(3)	202,4(1)	17293,7(9)	2103,8(3)	13,2(8)	730,2(6)	0,947(4)	3882,7(1)	0,372(2)	14,83(8)	24,9(9)
MU70 ^P	252,7(5)	14,2(8)	2465,9(7)	19,9(3)	418,5(5)	17040,2(6)	2628,0(8)	33,6(6)	742,0(2)	1,05(6)	3064,8(9)	11,17(6)	36,18(9)	39,3(5)
MU71 ^P	237,8(8)	12,2(6)	2464,9(9)	20,4(8)	414,8(1)	17479,8(1)	2717,6(5)	34,1(7)	783,7(1)	1,07(6)	3132,9(5)	8,18(8)	38,36(4)	40,3(2)
MU72 ^P	252,2(3)	14,4(4)	2515,7(4)	19,8(9)	416,3(8)	14964,3(9)	2710,1(4)	33,6(9)	739,0(4)	1,01(4)	3105,3(5)	10,53(4)	37,02(2)	39,4(4)

EX: Extremadura, MU: Murcia, a: agridulce, d: dulce, p: picante. Valor entre paréntesis: error asociado a la última cifra significativa

A1.2. Valores de ASTA, L*, a*, b*, C* y H en muestras de pimentón

Muestra	ASTA	L	a*	b	C	H
EX01 ^a	107	62,89	17,55	78,27	80,21	77,36
EX02 ^a	108	61,79	17,64	77,45	79,43	77,17
EX03 ^a	92	64,07	12,12	64,91	66,03	79,42
EX04 ^a	90	65,27	11,82	65,53	66,59	79,77
EX05 ^a	85	84,3	16,00	90,05	91,46	79,93
EX06 ^a	87	83,36	15,96	89,55	90,96	79,89
EX07 ^a	92	65,04	19,36	82,8	85,03	76,84
EX08 ^a	92	65,45	20,74	83,68	86,21	76,08
EX09 ^a	134	61,18	16,33	73,6	75,39	77,49
EX10 ^a	137	60,04	16,01	72,65	74,39	77,57
EX12 ^a	102	81,61	21,54	103,12	105,34	78,2
EX13 ^a	102	82,26	21,59	103,96	106,18	78,26
EX15 ^a	99	76,86	19,69	97,9	99,86	78,63
EX16 ^a	102	75,33	19,54	96,8	98,76	78,59
EX17 ^a	110	78,71	19,69	93,22	95,28	78,08
EX18 ^a	113	78,18	20,35	93,96	96,13	77,78
EX23 ^a	67	87,67	9,01	64,46	65,09	82,05
EX24 ^a	70	86,58	9,3	65,35	66,01	81,9
MU03 ^a	50	85,44	5,17	50,72	50,98	84,18
MU04 ^a	50	85,73	5,3	51,21	51,49	84,09
MU05 ^a	66	85,48	10,4	71,68	72,43	81,75
MU06 ^a	67	85,58	11,36	72,4	73,29	81,08
MU07 ^a	103	69,8	14,68	75,78	77,19	79,03
MU09 ^a	106	68,15	14,49	74,87	76,26	79,05
MU10 ^a	95	80,84	13,54	77,05	78,23	80,03
MU11 ^a	98	79,44	13,76	76,86	78,09	79,85
MU15 ^a	109	88,82	11,29	77,49	78,31	81,71
MU16 ^a	113	88,47	12,27	79,6	80,54	81,24
MU17 ^a	106	52,28	20,19	84,66	87,04	76,59
MU18 ^a	112	48,38	19,28	80,52	82,8	76,53
MU19 ^a	125	25,27	12,09	47,05	48,58	75,59
MU20 ^a	130	21,92	11,03	42,42	43,83	75,43
EX25 ^d	103	56,67	12,52	62,49	63,74	78,67
EX26 ^d	99	60,29	12,46	65,24	66,42	79,18
EX27 ^d	82	82,78	12,35	74,5	75,52	80,59

A1.2. Continuación

Muestra	ASTA	L*	a*	b*	C*	H
EX28 ^d	87	78,55	11,9	71,31	72,29	80,53
EX29 ^d	130	59,65	17,35	77,62	79,54	77,4
EX30 ^d	139	53,67	16,21	72,14	73,94	77,34
EX33 ^d	96	74,6	19,28	88,54	90,61	77,71
EX34 ^d	102	69,65	18,37	84,36	86,34	77,72
EX37 ^d	129	47,18	18,54	80,43	82,53	77,02
EX38 ^d	131	45,67	18,2	78,44	80,52	76,93
EX39 ^d	100	71,87	22,24	97,00	99,52	77,09
EX40 ^d	110	64,19	21,38	90,34	92,83	76,68
EX41 ^d	88	71,71	20,27	84,76	87,15	76,55
EX42 ^d	90	71,14	20,53	85,26	87,7	76,46
EX43 ^d	119	66,63	17,59	80,75	82,64	77,71
EX44 ^d	125	62,11	16,75	76,94	78,74	77,72
EX45 ^d	105	39,15	5,92	36,43	36,91	80,77
EX46 ^d	105	39,06	6,00	36,33	36,82	80,63
EX47 ^d	123	39,04	4,47	31,8	32,11	82
EX48 ^d	127	37,59	4,50	31,36	31,68	81,83
MU22 ^d	90	69,41	13,11	65,21	66,52	78,63
MU23 ^d	91	68,4	12,28	64,16	65,33	79,17
MU24 ^d	63	88,08	7,53	64,44	64,88	83,34
MU25 ^d	63	88,13	7,62	64,99	65,44	83,32
MU27 ^d	82	69,39	15,05	71,17	72,75	78,06
MU28 ^d	86	65,55	14,37	67,97	69,47	78,06
MU29 ^d	105	71,54	17,32	80,95	82,79	77,92
MU30 ^d	115	64,77	16,6	75,94	77,73	77,67
MU31 ^d	89	70,51	13,46	65,21	66,58	78,34
MU35 ^d	87	72,15	13,58	66,42	67,79	78,45
MU36 ^d	98	61,81	14,62	68,61	70,15	77,97
MU37 ^d	106	53,09	12,48	59,61	60,9	78,17
MU38 ^d	124	37,59	13,15	62,52	63,89	78,12
MU42 ^d	115	44,17	14,78	61,1	62,86	76,4
EX49 ^p	75	85,12	12,24	74,48	75,48	80,67
EX50 ^p	74	85,85	12,25	75,26	76,25	80,76
EX51 ^p	141	52,99	13,03	64,78	66,08	78,62
EX52 ^p	134	57,64	14,26	69,22	70,67	75,05

A1.2. Continuación

Muestra	ASTA	L*	a*	b*	C*	H
EX53 ^P	72	41,21	2,75	22,85	23,02	83,15
EX54 ^P	72	41,11	2,77	22,94	23,11	83,11
EX55 ^P	106	58,7	33,12	108,01	112,97	72,95
EX56 ^P	115	59,62	34,71	109,77	115,13	72,45
EX57 ^P	143	40,94	20,29	76,12	78,77	75,07
EX58 ^P	154	28,72	15,81	56,35	58,52	74,32
EX59 ^P	104	71,61	24,16	104,31	107,07	76,96
EX60 ^P	110	67,75	23,69	101,05	103,79	76,81
EX61 ^P	131	46,26	17,63	77,39	79,37	77,16
EX62 ^P	129	48,18	17,7	79,61	81,55	77,47
EX63 ^P	88	21,22	1,03	10,22	10,27	84,26
EX64 ^P	88	21,04	1,01	10,21	10,26	84,36
EX65 ^P	102	69,4	21,29	92,17	94,6	76,99
EX66 ^P	105	67,26	21,33	91,02	93,48	76,81
EX67 ^P	80	58,9	13,14	65,82	67,12	78,71
EX68 ^P	85	53,02	11,76	60,25	61,39	78,95
EX69 ^P	124	46,87	19,99	82,85	85,23	76,44
EX70 ^P	132	39,81	17,58	72,69	74,79	76,41
EX71 ^P	100	58,28	3,97	32,46	32,7	83,02
EX72 ^P	104	56,55	4,11	31,91	32,18	82,66
MI45 ^P	79	61,72	5,04	41,67	41,98	83,11
MU46 ^P	62	74,43	5,47	47,02	47,34	83,37
MU47 ^P	90	78,96	11,66	68,22	69,21	80,3
MU48 ^P	93	76,89	11,61	67,4	68,39	80,23
MU50 ^P	78	76,92	15,43	80,13	81,6	79,1
MU53 ^P	81	73,99	15,1	77,79	79,24	79,01
MU54 ^P	97	77,22	13,45	74,96	76,15	79,82
MU55 ^P	105	72,24	13,05	72,42	73,59	79,78
MU57 ^P	105	62,95	15,48	73,03	74,65	78,03
MU58 ^P	107	61,7	15,11	72,34	73,9	78,2
MU59 ^P	113	67,64	18,32	84,33	86,3	77,74
MU60 ^P	119	64,68	18,54	83,06	85,1	77,42
MU61 ^P	89	65,24	10,1	56,78	57,68	79,92
MU62 ^P	95	61,4	9,83	54,55	55,43	79,78
MU63 ^P	114	81,36	6,61	52,04	52,46	82,76

A1.2. Continuación

Muestra	ASTA	L*	a*	b*	C*	H
MU64 ^P	116	80,54	6,61	51,36	51,79	82,67
MU65 ^P	108	49,93	13,09	55,78	57,3	76,79
MU66 ^P	118	44,68	12,36	52,81	54,23	76,83
MU70 ^P	99	60,85	3,22	24,79	24,99	82,59
MU71 ^P	100	60,21	3,13	24,76	24,96	82,8

EX: Extremadura; MU: Murcia; a: agridulce, d:dulce, p: picante

A1.3. Concentración de C, DHC para pimentón picante en mg Kg⁻¹.
SHU: Unidades Scoville

Muestra	[C]	[DHC]	SHU
EX49 ^P	246,34(1)	129,82(6)	3966
EX50 ^P	208,33(8)	108,88(7)	3354
EX51 ^P	115,65(6)	70,13(1)	1862
EX52 ^P	139,98(7)	88,35(9)	2254
EX53 ^P	34,14(8)	22,54(3)	550
EX54 ^P	26,36(3)	17,25(3)	424
EX55 ^P	269,71(1)	155,02(6)	4342
EX56 ^P	191,34(8)	89,95(5)	3081
EX57 ^P	59,18(7)	28,88(8)	953
EX58 ^P	59,85(0)	25,36(9)	964
EX59 ^P	105,77(4)	60,34(0)	1703
EX60 ^p	102,59(6)	78,41(1)	1652
EX61 ^P	78,5(1)	49,95(9)	1264
EX62 ^P	32,58(4)	62,31(8)	525
EX63 ^P	229,37(5)	160,92(2)	3693
EX64 ^P	278,09(9)	256,12(1)	4477
EX65 ^P	144,72(6)	102,91(4)	2330
EX66 ^P	135,61(2)	95,62(4)	2183
EX67 ^P	120,46(1)	113,82(1)	1939
EX68 ^P	94,52(0)	29,65(8)	1522
EX69 ^P	92,66(8)	69,56(3)	1492
EX70 ^P	83,96(3)	98,95(9)	1352
EX71 ^P	57,96(1)	41,96(8)	933
EX72 ^P	70,81(2)	55,47(6)	1140
MU45 ^P	90,85(2)	27,49(5)	1463
MU46 ^P	97,6(4)	58,54(0)	1571
MU47 ^P	547,23(0)	299,02(9)	8810
MU48 ^P	494,47(8)	443,62(7)	7961
MU50 ^P	136,46(7)	82,78(6)	2197
MU53 ^P	129,07(3)	72,15(7)	2078
MU54 ^P	904,22(0)	396,64(0)	14558
MU55 ^P	919,86(8)	476,36(2)	14810
MU57 ^P	153,93(9)	94,45(5)	2478
MU58 ^P	133,41(3)	95,42(6)	2148
MU59 ^P	613,16(5)	360,21(6)	9872

A1.3. Continuación

Muestra	[C]	[DHC]	SHU
MU60 ^P	586,04(2)	401,32(1)	9435
MU61 ^P	243,28(4)	198,48(9)	3917
MU62 ^P	145,58(1)	165,44(3)	2344
MU63 ^P	238,26(0)	135,45(2)	3836
MU64 ^P	252,78(1)	118,06(6)	4070
MU65 ^P	150,25(5)	68,79(9)	2419
MU66 ^P	126,45(0)	110,12(2)	2036
MU70 ^P	154,09(0)	81,03(5)	2481
MU71 ^P	163,16(4)	76,40(9)	2627

EX: Extremadura, MU: Murcia, p: picante, C: capsaicina, DHC: dihidrocapsaicina, SHU: unidades Scoville.

Valor entre paréntesis: error asociado a la última cifra significativa

A1.4. Concentración de C, DHC para pimentón agridulces en mg Kg⁻¹.

SHU: Unidades Scoville

Muestra	[C]	[DHC]	SHU
EX01 ^a	3,59(4)	5,71(2)	58
EX02 ^a	3,95(9)	6,28(1)	64
EX03 ^a	7,33(6)	1,59(0)	118
EX04 ^a	7,87(1)	1,71(6)	127
EX05 ^a	4,7(7)	n.d	76
EX06 ^a	5,12(4)	n.d	82
EX07 ^a	n.d.	n.d	nd
EX08 ^a	n.d.	n.d	nd
EX09 ^a	n.d.	n.d	nd
EX10 ^a	n.d.	n.d	nd
EX12 ^a	20,14(6)	20,65(0)	324
EX13 ^a	18,73(7)	19,21(5)	302
EX15 ^a	6,17(9)	n.d	99
EX16 ^a	7,41(3)	n.d	119
EX17 ^a	1,28(6)	8,35(8)	21
EX18 ^a	1,54(2)	6,68(2)	25
EX23 ^a	6,29(3)	4,87(9)	101
EX24 ^a	6,18(1)	4,79(2)	99
MU03 ^a	9,78(0)	3,21(5)	157
MU04 ^a	8,53(6)	2,8(1)	137
MU05 ^a	11,36(3)	n.d	183
MU06 ^a	10,23(3)	n.d	165
MU07 ^a	2,8(6)	n.d	45
MU09 ^a	3,08(8)	n.d	49
MU10 ^a	14,74(6)	7,96(7)	237
MU11 ^a	12,57(7)	6,78(7)	202
MU15 ^a	3,83(0)	n.d	62
MU16 ^a	3,75(7)	n.d	60
MU17 ^a	6,23(5)	n.d	100
MU18 ^a	6,53(6)	n.d	105
MU19 ^a	7,93(5)	13,77(5)	128
MU20 ^a	7,14(2)	12,39(2)	115

EX: Extremadura, MU: Murcia, C: capsaicina, DHC: dihidrocapsaicina, SHU: unidades Scoville. Valor entre paréntesis: error asociado a la última

A1.5. Concentración de C, DHC para pimentón dulce en mg Kg⁻¹.

SHU: Unidades Scoville

Muestra	[C]	[DHC]	SHU
EX25 ^d	7,59(0)	3,41(9)	122
EX26 ^d	7,81(9)	2,39(5)	126
EX27 ^d	n.d.	n.d.	77
EX28 ^d	4,08(3)	2,6(0)	66
EX29 ^d	n.d.	n.d.	nd
EX30 ^d	3,47(7)	1,27(8)	56
EX33 ^d	3,54(6)	1,79(0)	57
EX34 ^d	3,51(4)	2,06(7)	57
EX37 ^d	n.d.	n.d.	nd
EX38 ^d	n.d.	n.d.	nd
EX39 ^d	n.d.	n.d.	nd
EX40 ^d	2,58(2)	n.d.	42
EX41 ^d	2,95(3)	1,34(8)	47
EX42 ^d	2,99(9)	n.d.	48
EX43 ^d	3,11(2)	1,76(2)	50
EX44 ^d	3,75(0)	n.d.	60
EX45 ^d	3,56(9)	2,78(0)	57
EX46 ^d	n.d.	n.d.	nd
EX47 ^d	5,22(4)	n.d.	84
EX48 ^d	4,42(9)	2,05(9)	71
MU22 ^d	n.d.	n.d.	nd
MU23 ^d	n.d.	n.d.	nd
MU24 ^d	3,96(9)	n.d.	64
MU25 ^d	2,94(0)	1,72(1)	47
MU27 ^d	8,94(2)	4,68(6)	144
MU28 ^d	5,48(0)	3,18(5)	88
MU29 ^d	n.d.	n.d.	nd
MU30 ^d	n.d.	n.d.	nd
MU31 ^d	5,72(5)	1,44(0)	92
MU35 ^d	4,97(2)	2,08(2)	80
MU36 ^d	5,41(4)	1,12(5)	87
MU37 ^d	4,47(3)	6,76(2)	72

EX: Extremadura, MU: Murcia, C: capsaicina, DHC: dihidrocapsaicina, SHU: unidades Scoville. Valor entre paréntesis: error asociado a la última cifra significativa

ANEXO 2

A2.1. Distribución del estadístico Lilliefors

v	α				
	0,2	0,15	0,05	0,01	0,001
4	0,303	0,321	0,376	0,413	0,433
5	0,289	0,303	0,343	0,397	0,439
6	0,269	0,281	0,323	0,371	0,424
7	0,252	0,264	0,304	0,351	0,402
8	0,239	0,25	0,288	0,333	0,384
9	0,227	0,238	0,274	0,317	0,365
10	0,217	0,228	0,262	0,304	0,352
11	0,208	0,218	0,251	0,291	0,338
12	0,2	0,210	0,242	0,281	0,325
13	0,193	0,202	0,234	0,271	0,314
14	0,187	0,196	0,226	0,262	0,305
15	0,181	0,190	0,219	0,254	0,296
16	0,176	0,184	0,213	0,247	0,287
17	0,171	0,179	0,207	0,24	0,279
18	0,167	0,175	0,202	0,234	0,273
19	0,163	0,170	0,197	0,228	0,266
20	0,159	0,166	0,192	0,223	0,26
25	0,143	0,150	0,173	0,201	0,236
30	0,131	0,138	0,159	0,185	0,217
>30	0,740	0,770	0,890	1,040	1,220
	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}

zA2.2. Valores de la distribución t de Student

v	α								
	0,25	0,2	0,1	0,025	0,01	0,05	0,001	0,005	0,0001
1	2,41	3,08	6,31	25,45	63,66	12,71	636,62	127,32	6366,20
2	1,60	1,89	2,92	6,21	9,92	4,30	31,60	14,09	99,99
3	1,42	1,64	2,35	4,18	5,84	3,18	12,92	7,45	28,00
4	1,34	1,53	2,13	3,50	4,60	2,78	8,61	5,60	15,54
5	1,30	1,48	2,02	3,16	4,03	2,57	6,87	4,77	11,18
6	1,27	1,44	1,94	2,97	3,71	2,45	5,96	4,32	9,08
7	1,25	1,41	1,89	2,84	3,50	2,36	5,41	4,03	7,88
8	1,24	1,40	1,86	2,75	3,36	2,31	5,04	3,83	7,12
9	1,23	1,38	1,83	2,69	3,25	2,26	4,78	3,69	6,59
10	1,22	1,37	1,81	2,63	3,17	2,23	4,59	3,58	6,21
11	1,21	1,36	1,80	2,59	3,11	2,20	4,44	3,50	5,92
12	1,21	1,36	1,78	2,56	3,05	2,18	4,32	3,43	5,69
13	1,20	1,35	1,77	2,53	3,01	2,16	4,22	3,37	5,51
14	1,20	1,35	1,76	2,51	2,98	2,14	4,14	3,33	5,36
15	1,20	1,34	1,75	2,49	2,95	2,13	4,07	3,29	5,24
16	1,19	1,34	1,75	2,47	2,92	2,12	4,01	3,25	5,13
17	1,19	1,33	1,74	2,46	2,90	2,11	3,97	3,22	5,04
18	1,19	1,33	1,73	2,45	2,88	2,10	3,92	3,20	4,97
19	1,19	1,33	1,73	2,43	2,86	2,09	3,88	3,17	4,90
20	1,18	1,33	1,72	2,42	2,85	2,09	3,85	3,15	4,84
21	1,18	1,32	1,72	2,41	2,83	2,08	3,82	3,14	4,78
22	1,18	1,32	1,72	2,41	2,82	2,07	3,79	3,12	4,74
23	1,18	1,32	1,71	2,40	2,81	2,07	3,77	3,10	4,69
24	1,18	1,32	1,71	2,39	2,80	2,06	3,75	3,09	4,65
25	1,18	1,32	1,71	2,38	2,79	2,06	3,73	3,08	4,62
26	1,18	1,31	1,71	2,38	2,78	2,06	3,71	3,07	4,59
27	1,18	1,31	1,70	2,37	2,77	2,05	3,69	3,06	4,56
28	1,17	1,31	1,70	2,37	2,76	2,05	3,67	3,05	4,53
29	1,17	1,31	1,70	2,36	2,76	2,05	3,66	3,04	4,51
30	1,17	1,31	1,70	2,36	2,75	2,04	3,65	3,03	4,48
31	1,17	1,31	1,70	2,36	2,74	2,04	3,63	3,02	4,46
32	1,17	1,31	1,69	2,35	2,74	2,04	3,62	3,01	4,44
33	1,17	1,31	1,69	2,35	2,73	2,03	3,61	3,01	4,42
34	1,17	1,31	1,69	2,35	2,73	2,03	3,60	3,00	4,41
35	1,17	1,31	1,69	2,34	2,72	2,03	3,59	3,00	4,39
40	1,17	1,30	1,68	2,33	2,70	2,02	3,55	2,97	4,32
50	1,16	1,30	1,68	2,31	2,68	2,01	3,50	2,94	4,23
75	1,16	1,29	1,67	2,29	2,64	1,99	3,43	2,89	4,11
∞	1,16	1,29	1,66	2,28	2,63	1,98	3,39	2,87	4,05

Los niveles de α son para dos colas. Si se aplica una prueba unidireccional, α sería $\alpha/2$

A2.3. Valores críticos F de Fisher para una cola. v_1 : numerador; v_2 : denominador

α	v_2	v_1							
		1	2	3	4	5	6	7	8
0,05	1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9
0,025		647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7
0,01		4052,2	4999,5	5403,4	5624,6	5763,6	5859,0	5928,4	5981,1
0,005		16210,7	19999,5	21614,7	22499,6	23055,8	23437,1	23714,6	23925,4
0,05	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4
0,025		38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4
0,01		98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4
0,005		198,5	199,0	199,2	199,2	199,3	199,3	199,4	199,4
0,05	3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85
0,025		17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54
0,01		34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49
0,005		55,55	49,80	47,47	46,19	45,39	44,84	44,43	44,13
0,05	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04
0,025		12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98
0,01		21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80
0,005		31,33	26,28	24,26	23,15	22,46	21,97	21,62	21,35
0,05	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82
0,025		10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76
0,01		16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29
0,005		22,78	18,31	16,53	15,56	14,94	14,51	14,20	13,96
0,05	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15
0,025		8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60
0,01		13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10
0,005		18,63	14,54	12,92	12,03	11,46	11,07	10,79	10,57
0,05	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73
0,025		8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90
0,01		12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84
0,005		16,24	12,40	10,88	10,05	9,52	9,16	8,89	8,68
0,05	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44
0,025		7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43
0,01		11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03
0,005		14,69	11,04	9,60	8,81	8,30	7,95	7,69	7,50
0,05	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23
0,025		7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10
0,01		10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47
0,005		13,61	10,11	8,72	7,96	7,47	7,13	6,88	6,69
0,05	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07
0,025		6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85
0,01		10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06
0,005		12,83	9,43	8,08	7,34	6,87	6,54	6,30	6,12

A2.3. Continuación

α	v_2	v_1							
		1	2	3	4	5	6	7	8
0,05	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95
0,025		6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66
0,01		9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74
0,005		12,23	8,91	7,60	6,88	6,42	6,10	5,86	5,68
0,05	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85
0,025		6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51
0,01		9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50
0,005		11,75	8,51	7,23	6,52	6,07	5,76	5,52	5,35
0,05	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64
0,025		6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20
0,01		8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00
0,005		10,80	7,70	6,48	5,80	5,37	5,07	4,85	4,67
0,05	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45
0,025		5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91
0,01		8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56
0,005		9,94	6,99	5,82	5,17	4,76	4,47	4,26	4,09
0,05	50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13
0,025		5,34	3,97	3,39	3,05	2,83	2,67	2,55	2,46
0,01		7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89
0,005		8,63	5,90	4,83	4,23	3,85	3,58	3,38	3,22
0,05	75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06
0,025		5,23	3,88	3,30	2,96	2,74	2,58	2,46	2,37
0,01		6,99	4,90	4,05	3,58	3,27	3,05	2,89	2,76
0,005		8,37	5,69	4,63	4,05	3,67	3,41	3,21	3,05
0,05	100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03
0,025		5,18	3,83	3,25	2,92	2,70	2,54	2,42	2,32
0,01		6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69
0,005		8,24	5,59	4,54	3,96	3,59	3,33	3,13	2,97
0,05	1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95
0,025		5,04	3,70	3,13	2,80	2,58	2,42	2,30	2,20
0,01		6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53
0,005		7,91	5,33	4,30	3,74	3,37	3,11	2,92	2,77

A2.3. Continuación

α	v_2	v_1							
		10	11	12	15	20	50	75	1000
0,05	1	241,9	243,0	243,9	245,9	248,0	251,8	252,6	254,2
0,025		968,6	973,0	976,7	984,9	993,1	1008,1	1011,5	1017,7
0,01		6055,8	6083,3	6106,3	6157,3	6208,7	6302,5	6323,6	6362,7
0,005		24224,5	24334,4	24426,4	24630,2	24836,0	25211,1	25295,3	25451,7
0,05	2	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5
0,025		39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5
0,01		99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5
0,005		199,4	199,4	199,4	199,4	199,4	199,5	199,5	199,5
0,05	3	8,79	8,76	8,74	8,70	8,66	8,58	8,56	8,53
0,025		14,42	14,37	14,34	14,25	14,17	14,01	13,97	13,91
0,01		27,23	27,13	27,05	26,87	26,69	26,35	26,28	26,14
0,005		43,69	43,52	43,39	43,08	42,78	42,21	42,09	41,85
0,05	4	5,96	5,94	5,91	5,86	5,80	5,70	5,68	5,63
0,025		8,84	8,79	8,75	8,66	8,56	8,38	8,34	8,26
0,01		14,55	14,45	14,37	14,20	14,02	13,69	13,61	13,47
0,005		20,97	20,82	20,70	20,44	20,17	19,67	19,55	19,34
0,05	5	4,74	4,70	4,68	4,62	4,56	4,44	4,42	4,37
0,025		6,62	6,57	6,52	6,43	6,33	6,14	6,10	6,02
0,01		10,05	9,96	9,89	9,72	9,55	9,24	9,17	9,03
0,005		13,62	13,49	13,38	13,15	12,90	12,45	12,35	12,16
0,05	6	4,06	4,03	4,00	3,94	3,87	3,75	3,73	3,67
0,025		5,46	5,41	5,37	5,27	5,17	4,98	4,94	4,86
0,01		7,87	7,79	7,72	7,56	7,40	7,09	7,02	6,89
0,005		10,25	10,13	10,03	9,81	9,59	9,17	9,07	8,89
0,05	7	3,64	3,60	3,57	3,51	3,44	3,32	3,29	3,23
0,025		4,76	4,71	4,67	4,57	4,47	4,28	4,23	4,15
0,01		6,62	6,54	6,47	6,31	6,16	5,86	5,79	5,66
0,005		8,38	8,27	8,18	7,97	7,75	7,35	7,26	7,09
0,05	8	3,35	3,31	3,28	3,22	3,15	3,02	2,99	2,93
0,025		4,30	4,24	4,20	4,10	4,00	3,81	3,76	3,68
0,01		5,81	5,73	5,67	5,52	5,36	5,07	5,00	4,87
0,005		7,21	7,10	7,01	6,81	6,61	6,22	6,13	5,96
0,05	9	3,14	3,10	3,07	3,01	2,94	2,80	2,77	2,71
0,025		3,96	3,91	3,87	3,77	3,67	3,47	3,43	3,34
0,01		5,26	5,18	5,11	4,96	4,81	4,52	4,45	4,32
0,005		6,42	6,31	6,23	6,03	5,83	5,45	5,37	5,20
0,05	10	2,98	2,94	2,91	2,85	2,77	2,64	2,60	2,54
0,025		3,72	3,66	3,62	3,52	3,42	3,22	3,18	3,09
0,01		4,85	4,77	4,71	4,56	4,41	4,12	4,05	3,92
0,005		5,85	5,75	5,66	5,47	5,27	4,90	4,82	4,65

A2.3. Continuación

α	v_2	v_1							
		10	11	12	15	20	50	75	1000
0,05	11	2,85	2,82	2,79	2,72	2,65	2,51	2,47	2,41
0,025		3,53	3,47	3,43	3,33	3,23	3,03	2,98	2,89
0,01		4,54	4,46	4,40	4,25	4,10	3,81	3,74	3,61
0,005		5,42	5,32	5,24	5,05	4,86	4,49	4,40	4,24
0,05	12	2,75	2,72	2,69	2,62	2,54	2,40	2,37	2,30
0,025		3,37	3,32	3,28	3,18	3,07	2,87	2,82	2,73
0,01		4,30	4,22	4,16	4,01	3,86	3,57	3,50	3,37
0,005		5,09	4,99	4,91	4,72	4,53	4,17	4,08	3,92
0,05	15	2,54	2,51	2,48	2,40	2,33	2,18	2,14	2,07
0,025		3,06	3,01	2,96	2,86	2,76	2,55	2,50	2,40
0,01		3,80	3,73	3,67	3,52	3,37	3,08	3,01	2,88
0,005		4,42	4,33	4,25	4,07	3,88	3,52	3,44	3,27
0,05	20	2,35	2,31	2,28	2,20	2,12	1,97	1,93	1,85
0,025		2,77	2,72	2,68	2,57	2,46	2,25	2,20	2,09
0,01		3,37	3,29	3,23	3,09	2,94	2,64	2,57	2,43
0,005		3,85	3,76	3,68	3,50	3,32	2,96	2,87	2,70
0,05	50	2,03	1,99	1,95	1,87	1,78	1,60	1,55	1,45
0,025		2,32	2,26	2,22	2,11	1,99	1,75	1,69	1,56
0,01		2,70	2,63	2,56	2,42	2,27	1,95	1,87	1,70
0,005		2,99	2,90	2,82	2,65	2,47	2,10	2,00	1,80
0,05	75	1,96	1,92	1,88	1,80	1,71	1,52	1,47	1,35
0,025		2,22	2,17	2,12	2,01	1,90	1,65	1,58	1,43
0,01		2,57	2,49	2,43	2,29	2,13	1,81	1,72	1,53
0,005		2,82	2,74	2,66	2,49	2,31	1,92	1,82	1,61
0,05	1000	1,84	1,80	1,76	1,68	1,58	1,36	1,30	1,11
0,025		2,06	2,01	1,96	1,85	1,72	1,45	1,36	1,13
0,01		2,34	2,27	2,20	2,06	1,90	1,54	1,44	1,16
0,005		2,54	2,45	2,38	2,21	2,02	1,61	1,50	1,18

A2.4. Valores críticos F de Fisher para dos colas. v_1 : numerador; v_2 : denominador

α	v_2	v_1							
		1	2	3	4	5	6	7	8
0,05	1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7
0,025		2593,2	3199,5	3458,0	3599,6	3688,6	3749,6	3794,0	3827,7
0,01		16210,7	19999,5	21614,7	22499,6	23055,8	23437,1	23714,6	23925,4
0,005		64844,9	79999,5	86460,3	89999,6	92224,4	93749,6	94859,4	95702,7
0,05	2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4
0,025		38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4
0,01		98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4
0,005		398,5	399,0	399,2	399,2	399,3	399,3	399,4	399,4
0,05	3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54
0,025		17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54
0,01		55,55	49,80	47,47	46,19	45,39	44,84	44,43	44,13
0,005		89,58	79,93	76,06	73,95	72,62	71,71	71,04	70,53
0,05	4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98
0,025		18,62	15,89	14,77	14,15	13,75	13,48	13,28	13,13
0,01		31,33	26,28	24,26	23,15	22,46	21,97	21,62	21,35
0,005		45,67	38,00	34,96	33,30	32,26	31,54	31,02	30,62
0,05	5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76
0,025		14,52	11,93	10,86	10,28	9,90	9,64	9,45	9,31
0,01		22,78	18,31	16,53	15,56	14,94	14,51	14,20	13,96
0,005		31,41	24,96	22,43	21,05	20,18	19,58	19,14	18,80
0,05	6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60
0,025		12,40	9,93	8,91	8,35	8,00	7,75	7,56	7,42
0,01		18,63	14,54	12,92	12,03	11,46	11,07	10,79	10,57
0,005		24,81	19,10	16,87	15,65	14,88	14,35	13,96	13,67
0,05	7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90
0,025		11,12	8,74	7,77	7,22	6,88	6,64	6,46	6,32
0,01		16,24	12,40	10,88	10,05	9,52	9,16	8,89	8,68
0,005		21,11	15,89	13,84	12,73	12,03	11,55	11,19	10,91
0,05	8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43
0,025		10,28	7,96	7,02	6,49	6,15	5,92	5,74	5,61
0,01		14,69	11,04	9,60	8,81	8,30	7,95	7,69	7,50
0,005		18,78	13,89	11,98	10,94	10,28	9,83	9,49	9,24
0,05	9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10
0,025		9,68	7,42	6,49	5,98	5,65	5,41	5,24	5,11
0,01		13,61	10,11	8,72	7,96	7,47	7,13	6,88	6,69
0,005		17,19	12,54	10,73	9,74	9,12	8,68	8,36	8,12
0,05	10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85
0,025		9,23	7,01	6,10	5,60	5,27	5,04	4,88	4,74
0,01		12,83	9,43	8,08	7,34	6,87	6,54	6,30	6,12
0,005		16,04	11,57	9,83	8,89	8,29	7,87	7,56	7,33

A2.4. Continuación

α	v_2	v_1							
		1	2	3	4	5	6	7	8
0,05	11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66
0,025		8,89	6,70	5,81	5,31	4,99	4,76	4,60	4,47
0,01		12,23	8,91	7,60	6,88	6,42	6,10	5,86	5,68
0,005		15,17	10,85	9,17	8,25	7,67	7,27	6,97	6,74
0,05	12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51
0,025		8,61	6,45	5,57	5,08	4,76	4,54	4,37	4,25
0,01		11,75	8,51	7,23	6,52	6,07	5,76	5,52	5,35
0,005		14,49	10,29	8,65	7,76	7,20	6,80	6,51	6,29
0,05	15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20
0,025		8,05	5,95	5,10	4,62	4,31	4,09	3,93	3,80
0,01		10,80	7,70	6,48	5,80	5,37	5,07	4,85	4,67
0,005		13,13	9,17	7,63	6,80	6,26	5,89	5,62	5,40
0,05	20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91
0,025		7,53	5,50	4,67	4,20	3,90	3,69	3,53	3,40
0,01		9,94	6,99	5,82	5,17	4,76	4,47	4,26	4,09
0,005		11,94	8,21	6,76	5,97	5,46	5,11	4,85	4,65
0,05	50	5,34	3,97	3,39	3,05	2,83	2,67	2,55	2,46
0,025		6,71	4,79	4,00	3,56	3,27	3,06	2,91	2,78
0,01		8,63	5,90	4,83	4,23	3,85	3,58	3,38	3,22
0,005		10,14	6,77	5,47	4,75	4,30	3,98	3,74	3,55
0,05	75	5,23	3,88	3,30	2,96	2,74	2,58	2,46	2,37
0,025		6,55	4,65	3,87	3,43	3,14	2,94	2,78	2,66
0,01		8,37	5,69	4,63	4,05	3,67	3,41	3,21	3,05
0,005		9,79	6,50	5,22	4,52	4,08	3,76	3,53	3,35
0,05	∞	5,04	3,70	3,13	2,80	2,58	2,42	2,30	2,20
0,025		6,26	4,40	3,64	3,21	2,93	2,72	2,57	2,45
0,01		7,91	5,33	4,30	3,74	3,37	3,11	2,92	2,77
0,005		9,19	6,03	4,81	4,14	3,71	3,40	3,18	3,00

A2.4. Continuación

α	v_2	v_1							
		10	11	12	15	20	50	75	∞
0,05	1	968,6	973,0	976,7	984,9	993,1	1008,1	1011,5	254,2
0,025		3875,6	3893,2	3907,9	3940,5	3973,5	4033,5	4047,0	4072,0
0,01		24224,5	24334,4	24426,4	24630,2	24836,0	25211,1	25295,3	25451,7
0,005		96899,0	97338,5	97706,5	98521,9	99344,9	100845,4	101182,0	101807,9
0,05	2	39,4	39,4	39,4	39,4	39,4	39,5	39,5	19,5
0,025		39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5
0,01		99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5
0,005		399,4	399,4	399,4	399,4	399,4	399,5	399,5	399,5
0,05	3	14,42	14,37	14,34	14,25	14,17	14,01	13,97	8,53
0,025		14,42	14,37	14,34	14,25	14,17	14,01	13,97	13,91
0,01		43,69	43,52	43,39	43,08	42,78	42,21	42,09	41,85
0,005		69,81	69,54	69,32	68,82	68,31	67,39	67,18	66,79
0,05	4	8,84	8,79	8,75	8,66	8,56	8,38	8,34	5,63
0,025		12,91	12,83	12,76	12,61	12,45	12,17	12,10	11,98
0,01		20,97	20,82	20,70	20,44	20,17	19,67	19,55	19,34
0,005		30,04	29,83	29,66	29,26	28,86	28,12	27,95	27,64
0,05	5	6,62	6,57	6,52	6,43	6,33	6,14	6,10	4,37
0,025		9,10	9,02	8,95	8,81	8,66	8,38	8,31	8,19
0,01		13,62	13,49	13,38	13,15	12,90	12,45	12,35	12,16
0,005		18,32	18,14	17,99	17,66	17,32	16,69	16,55	16,28
0,05	6	5,46	5,41	5,37	5,27	5,17	4,98	4,94	3,67
0,025		7,22	7,14	7,08	6,94	6,79	6,52	6,45	6,34
0,01		10,25	10,13	10,03	9,81	9,59	9,17	9,07	8,89
0,005		13,24	13,08	12,95	12,65	12,35	11,78	11,66	11,41
0,05	7	4,76	4,71	4,67	4,57	4,47	4,28	4,23	4,15
0,025		6,12	6,05	5,99	5,85	5,70	5,43	5,37	5,26
0,01		8,38	8,27	8,18	7,97	7,75	7,35	7,26	7,09
0,005		10,52	10,38	10,25	9,98	9,70	9,17	9,05	8,83
0,05	8	4,30	4,24	4,20	4,10	4,00	3,81	3,76	2,93
0,025		5,41	5,34	5,28	5,14	5,00	4,73	4,67	4,55
0,01		7,21	7,10	7,01	6,81	6,61	6,22	6,13	5,96
0,005		8,87	8,73	8,61	8,35	8,09	7,59	7,48	7,26
0,05	9	3,96	3,91	3,87	3,77	3,67	3,47	3,43	3,34
0,025		4,92	4,85	4,79	4,65	4,51	4,24	4,18	4,06
0,01		6,42	6,31	6,23	6,03	5,83	5,45	5,37	5,20
0,005		7,77	7,63	7,52	7,28	7,02	6,54	6,43	6,22
0,05	10	3,72	3,66	3,62	3,52	3,42	3,22	3,18	2,54
0,025		4,56	4,48	4,42	4,29	4,15	3,88	3,82	3,70
0,01		5,85	5,75	5,66	5,47	5,27	4,90	4,82	4,65
0,005		6,99	6,86	6,75	6,51	6,27	5,80	5,69	5,49

A2.4. Continuación

α	v_2	v_1							
		10	11	12	15	20	50	75	∞
0,05	11	3,53	3,47	3,43	3,33	3,23	3,03	2,98	2,41
0,025		4,28	4,21	4,15	4,01	3,87	3,61	3,55	3,43
0,01		5,42	5,32	5,24	5,05	4,86	4,49	4,40	4,24
0,005		6,41	6,29	6,18	5,95	5,71	5,25	5,15	4,94
0,05	12	3,37	3,32	3,28	3,18	3,07	2,87	2,82	2,30
0,025		4,06	3,99	3,93	3,80	3,66	3,39	3,33	3,21
0,01		5,09	4,99	4,91	4,72	4,53	4,17	4,08	3,92
0,005		5,97	5,85	5,74	5,52	5,28	4,83	4,73	4,53
0,05	15	3,06	3,01	2,96	2,86	2,76	2,55	2,50	2,07
0,025		3,62	3,55	3,49	3,36	3,22	2,95	2,88	2,76
0,01		4,42	4,33	4,25	4,07	3,88	3,52	3,44	3,27
0,005		5,10	4,98	4,88	4,67	4,44	4,00	3,90	3,70
0,05	20	2,77	2,72	2,68	2,57	2,46	2,25	2,20	2,09
0,025		3,22	3,15	3,09	2,96	2,82	2,54	2,48	2,35
0,01		3,85	3,76	3,68	3,50	3,32	2,96	2,87	2,70
0,005		4,35	4,24	4,15	3,94	3,72	3,29	3,19	2,99
0,05	50	2,32	2,26	2,22	2,11	1,99	1,75	1,75	1,56
0,025		2,61	2,54	2,48	2,34	2,20	1,96	1,90	1,66
0,01		2,99	2,90	2,82	2,65	2,47	2,16	2,10	1,80
0,005		3,28	3,18	3,09	2,89	2,68	2,32	2,24	1,91
0,05	75	2,22	2,17	2,12	2,01	1,90	1,69	1,65	1,35
0,025		2,48	2,42	2,36	2,22	2,07	1,82	1,77	1,51
0,01		2,82	2,74	2,66	2,49	2,31	1,99	1,92	1,61
0,005		3,08	2,98	2,89	2,69	2,48	2,12	2,04	1,68
0,05	∞	2,06	2,01	1,96	1,85	1,72	1,45	1,36	1,13
0,025		2,27	2,20	2,14	2,01	1,86	1,52	1,42	1,15
0,01		2,54	2,45	2,38	2,21	2,02	1,61	1,50	1,18
0,005		2,74	2,64	2,55	2,36	2,14	1,68	1,55	1,19

A2.5. Valores críticos para la prueba de Tukey.

α	v_2	v_1																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0,05	1	18	29,98	32,82	37,08	40,41	43,12	45,4	47,36	49,07	50,59	51,96	53,2	54,33	55,36	56,32	57,22	58,04	58,83	59,56
0,01		90,03	135	164,3	185,6	202,2	215,8	227,2	237	245,6	253,2	260	266,2	271,8	277	281,8	286,3	290,4	294,3	298
0,05	2	6,1	8,33	9,8	10,88	11,74	12,44	13,03	13,54	13,99	14,39	14,75	15,08	15,38	15,65	15,91	16,14	16,37	16,57	16,77
0,01		14,04	19,02	22,29	24,72	26,63	28,2	29,53	30,68	31,69	32,59	33,4	34,13	34,81	35,43	36	36,53	37,03	37,5	37,95
0,05	3	4,5	5,91	6,82	7,5	8,04	8,48	8,85	9,18	9,46	9,72	9,95	10,15	10,35	10,53	10,69	10,84	10,98	11,11	11,24
0,01		8,26	10,62	12,17	13,33	14,24	15	15,64	16,2	16,69	17,13	17,53	17,89	18,22	18,52	18,81	19,07	19,32	19,55	19,77
0,05	4	3,93	5,04	5,76	6,29	6,71	7,05	7,34	7,6	7,83	8,03	8,21	8,37	8,52	8,66	8,79	8,91	9,03	9,13	9,23
0,01		6,51	8,12	9,17	9,96	10,58	11,1	11,55	11,93	12,27	12,57	12,84	13,09	13,32	13,53	13,73	13,91	14,08	14,24	14,4
0,05	5	3,64	4,6	5,22	5,67	6,03	6,33	6,58	6,8	6,99	7,17	7,32	7,47	7,6	7,72	7,83	7,93	8,03	8,12	8,21
0,01		5,7	6,97	7,8	8,42	8,91	9,32	9,67	9,97	10,24	10,48	10,7	10,89	11,08	11,24	11,4	11,55	11,68	11,81	11,93
0,05	6	3,46	4,34	4,9	5,31	5,63	5,89	6,12	6,32	6,49	6,65	6,79	6,92	7,03	7,14	7,24	7,34	7,43	7,51	7,59
0,01		5,24	6,33	7,03	7,56	7,97	8,32	8,61	8,87	9,1	9,3	9,49	9,65	9,81	9,95	10,08	10,21	10,32	10,43	10,54
0,05	7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6	6,16	6,3	6,43	6,55	6,66	6,76	6,85	6,94	7,02	7,09	7,17
0,01		4,95	5,92	6,54	7,01	7,37	7,68	7,94	8,17	8,37	8,55	8,71	8,86	9	9,12	9,24	9,35	9,46	9,55	9,65
0,05	8	3,26	4,04	4,53	4,89	5,17	5,4	5,6	5,77	5,92	6,05	6,18	6,29	6,39	6,48	6,57	6,65	6,73	6,8	6,87
0,01		4,74	5,63	6,2	6,63	6,96	7,24	7,47	7,68	7,87	8,03	8,18	8,31	8,44	8,55	8,66	8,76	8,85	8,94	9,03
0,05	9	3,2	3,95	4,42	4,76	5,02	5,24	5,43	5,6	5,74	5,87	5,98	6,09	6,19	6,28	6,36	6,44	6,51	6,58	6,64
0,01		4,6	5,43	5,96	6,35	6,66	6,91	7,13	7,32	7,49	7,65	7,78	7,91	8,03	8,13	8,23	8,32	8,41	8,49	8,57
0,05	10	3,15	3,88	4,33	4,65	4,91	5,12	5,3	5,46	5,6	5,72	5,83	5,93	6,03	6,11	6,2	6,27	6,34	6,4	6,47
0,01		4,48	5,27	5,77	6,14	6,43	6,67	6,87	7,05	7,21	7,36	7,48	7,6	7,71	7,81	7,91	7,99	8,07	8,15	8,22
0,05	11	3,11	3,82	4,26	4,57	4,82	5,03	5,2	5,35	5,49	5,61	5,71	5,81	5,9	5,99	6,06	6,14	6,2	6,26	6,33
0,01		4,39	5,27	5,62	5,97	6,25	6,48	6,67	6,84	6,99	7,13	7,25	7,36	7,46	7,56	7,65	7,73	7,81	7,88	7,95
0,05	12	3,08	3,77	4,2	4,51	4,75	4,95	5,12	5,27	5,4	5,51	5,62	5,71	5,8	5,88	5,95	6,03	6,09	6,15	6,21
0,01		4,32	5,04	5,5	5,84	6,1	6,32	6,51	6,67	6,81	6,94	7,06	7,17	7,26	7,36	7,44	7,52	7,59	7,66	7,73

A2.5. Continuación

α	v_2	v_1																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0,05	15	3,01	3,67	4,08	4,37	4,6	4,78	4,94	5,08	5,2	5,31	5,4	5,49	5,58	5,65	5,72	5,79	5,85	5,9	5,96
0,01		4,17	4,83	5,25	5,56	5,8	5,99	6,16	6,31	6,44	6,55	6,66	6,76	6,84	6,93	7	7,07	7,14	7,2	7,26
0,05	16	3	3,65	4,05	4,33	4,56	4,74	4,9	5,03	5,15	5,26	5,35	5,44	5,52	5,59	5,66	5,72	5,79	5,84	5,9
0,01		4,13	4,78	5,19	5,49	5,72	5,92	6,08	6,22	6,35	6,46	6,56	6,66	6,74	6,82	6,9	6,97	7,03	7,09	7,15
0,05	17	2,98	3,63	4,02	4,3	4,52	4,71	4,86	4,99	5,11	5,21	5,31	5,39	5,47	5,55	5,61	5,68	5,74	5,79	5,84
0,01		4,1	4,74	5,14	5,43	5,66	5,85	6,01	6,15	6,27	6,38	6,48	6,57	6,66	6,73	6,8	6,87	6,94	7	7,05
0,05	18	2,97	3,61	4	4,28	4,49	4,67	4,82	4,96	5,07	5,17	5,27	5,35	5,43	5,5	5,57	5,63	5,69	5,74	5,79
0,01		4,07	4,7	5,09	5,38	5,6	5,79	5,94	6,08	6,2	6,31	6,41	6,5	6,58	6,65	6,72	6,79	6,85	6,91	6,96
0,05	19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23	5,32	5,39	5,46	5,53	5,59	5,65	5,7	5,75
0,01		4,05	4,67	5,05	5,33	5,55	5,73	5,89	6,02	6,14	6,25	6,34	6,43	6,51	6,58	6,65	6,72	6,78	6,84	6,89
0,05	20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,9	5,01	5,11	5,2	5,28	5,36	5,43	5,49	5,55	5,61	5,66	5,71
0,01		4,02	4,64	5,02	5,29	5,51	5,69	5,84	5,97	6,09	6,19	6,29	6,37	6,45	6,52	6,59	6,65	6,71	6,76	6,82
0,05	30	2,89	3,49	3,84	4,1	4,3	4,46	4,6	4,72	4,83	4,92	5	5,08	5,15	5,21	5,27	5,33	5,38	5,43	5,48
0,01		3,89	4,45	4,8	5,05	5,24	5,4	5,54	5,65	5,76	5,85	5,93	6,01	6,08	6,14	6,2	6,26	6,31	6,36	6,41
0,05	40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,74	4,82	4,91	4,98	5,05	5,11	5,16	5,22	5,27	5,31	5,36
0,01		3,82	4,37	4,7	4,93	5,11	5,27	5,39	5,5	5,6	5,69	5,77	5,84	5,9	5,96	6,02	6,07	6,12	6,17	6,21
0,05	60	2,83	3,4	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81	4,88	4,94	5	5,06	5,11	5,16	5,2	5,24
0,01		3,76	4,28	4,6	4,82	4,99	5,13	5,25	5,36	5,45	5,53	5,6	5,67	5,73	5,79	5,84	5,89	5,93	5,98	6,02
0,05	120	2,8	3,36	3,69	3,92	4,1	4,24	4,36	4,48	4,56	4,64	4,72	4,78	4,84	4,9	4,95	5	5,05	5,09	5,13
0,01		3,7	4,1	4,5	4,71	4,87	5,01	5,12	5,21	5,3	5,38	5,44	5,51	5,56	5,61	5,66	5,71	5,75	5,79	5,83
0,05	∞	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62	4,68	4,74	4,8	4,85	4,89	4,93	4,97	5,01
0,01		3,64	4,12	4,4	4,6	4,76	4,88	4,99	5,08	5,16	5,23	5,29	5,35	5,4	5,45	5,49	5,54	5,57	5,61	5,65

A2.6. Distribución estándar normal. Valores correspondientes al área izquierda de z

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916

A2.6. Continuación

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

A2.7. Valores críticos $\chi^2_{(\alpha, \nu)}$ de la distribución χ^2_ν Chi-cuadrada

v	α								
	0,995	0,99	0,975	0,95	0,9	0,1	0,05	0,01	0,005
1	7,879	6,635	5,024	3,841	2,706	0,016	0,004	0,000	0,000
2	10,597	9,210	7,378	5,991	4,605	0,211	0,103	0,020	0,010
3	12,838	11,345	9,348	7,815	6,251	0,584	0,352	0,115	0,072
4	14,860	13,277	11,143	9,488	7,779	1,064	0,711	0,297	0,207
5	16,750	15,086	12,833	11,070	9,236	1,610	1,145	0,554	0,412
6	18,548	16,812	14,449	12,592	10,645	2,204	1,635	0,872	0,676
7	20,278	18,475	16,013	14,067	12,017	2,833	2,167	1,239	0,989
8	21,955	20,090	17,535	15,507	13,362	3,490	2,733	1,646	1,344
9	23,589	21,666	19,023	16,919	14,684	4,168	3,325	2,088	1,735
10	25,188	23,209	20,483	18,307	15,987	4,865	3,940	2,558	2,156
11	26,757	24,725	21,920	19,675	17,275	5,578	4,575	3,053	2,603
12	28,300	26,217	23,337	21,026	18,549	6,304	5,226	3,571	3,074
13	29,819	27,688	24,736	22,362	19,812	7,042	5,892	4,107	3,565
14	31,319	29,141	26,119	23,685	21,064	7,790	6,571	4,660	4,075
15	32,801	30,578	27,488	24,996	22,307	8,547	7,261	5,229	4,601
16	34,267	32,000	28,845	26,296	23,542	9,312	7,962	5,812	5,142
17	35,718	33,409	30,191	27,587	24,769	10,085	8,672	6,408	5,697
18	37,156	34,805	31,526	28,869	25,989	10,865	9,390	7,015	6,265
19	38,582	36,191	32,852	30,144	27,204	11,651	10,117	7,633	6,844
20	39,997	37,566	34,170	31,410	28,412	12,443	10,851	8,260	7,434
21	41,401	38,932	35,479	32,671	29,615	13,240	11,591	8,897	8,034
22	42,796	40,289	36,781	33,924	30,813	14,041	12,338	9,542	8,643
23	44,181	41,638	38,076	35,172	32,007	14,848	13,091	10,196	9,260
24	45,559	42,980	39,364	36,415	33,196	15,659	13,848	10,856	9,886
25	46,928	44,314	40,646	37,652	34,382	16,473	14,611	11,524	10,520
26	48,290	45,642	41,923	38,885	35,563	17,292	15,379	12,198	11,160
27	49,645	46,963	43,195	40,113	36,741	18,114	16,151	12,879	11,808
28	50,993	48,278	44,461	41,337	37,916	18,939	16,928	13,565	12,461
29	52,336	49,588	45,722	42,557	39,087	19,768	17,708	14,256	13,121
30	53,672	50,892	46,979	43,773	40,256	20,599	18,493	14,953	13,787
40	66,766	63,691	59,342	55,758	51,805	29,051	26,509	22,164	20,707
50	79,490	76,154	71,420	67,505	63,167	37,689	34,764	29,707	27,991
75	110,286	106,393	100,839	96,217	91,061	59,795	56,054	49,475	47,206
100	140,169	135,807	129,561	124,342	118,498	82,358	77,929	70,065	67,328

A2.8. Valores críticos de Grubbs

n	$\alpha=0,05$			$\alpha=0,01$		
	G(1)	G(2)	G(3)	G(1)	G(2)	G(3)
3	1.153	2		1.155	2	
4	1.463	2.43	0.9992	1.492	2.44	1
5	1.672	2.75	0.9817	1.749	2.8	0.9965
6	1.822	3.01	0.9436	1.944	3.1	0.9814
7	1.938	3.22	0.898	2.097	3.34	0.956
8	2.032	3.4	0.8522	2.221	3.54	0.925
9	2.11	3.55	0.8091	2.323	3.72	0.8918
10	2.176	3.68	0.7695	2.41	3.88	0.8586
12	2.285	3.91	0.7004	2.55	4.13	0.7957
13	2.331	4	0.6705	2.607	4.24	0.7667
15	2.409	4.17	0.6182	2.705	4.43	0.7141
20	2.557	4.49	0.5196	2.884	4.79	0.6091
25	2.663	4.73	0.4505	3.009	5.03	0.532
30	2.745	4.89	0.3992	3.103	5.19	0.4732
35	2.811	5.026	0.3595	3.178	5.326	0.427
40	2.866	5.15	0.3276	3.24	5.45	0.3896
50	2.956	5.35	0.2797	3.336	5.65	0.3328
60	3.025	5.5	0.245	3.411	5.8	0.2914
70	3.082	5.638	0.2187	3.471	5.938	0.2599
80	3.13	5.73	0.1979	3.521	6.03	0.235
90	3.171	5.82	0.181	3.563	6.12	0.2147
100	3.207	5.9	0.1671	3.6	6.2	0.198
110	3.239	5.968	0.1553	3.632	6.268	0.1838
120	3.267	6.03	0.1452	3.662	6.33	0.1716
130	3.294	6.086	0.1364	3.688	6.386	0.1611
140	3.318	6.137	0.1288	3.712	6.437	0.1519

A2.9. Valores del estadístico C de Cochran.

α	Grupos	Repeticiones								
		2	3	4	5	6	7	8	9	10
0,05	2	0.998	0.975	0.939	0.906	0.877	0.853	0.833	0.816	0.801
0,01		1.000	0.995	0.979	0.959	0.937	0.917	0.899	0.882	0.867
0,05	3	0.967	0.871	0.798	0.746	0.707	0.677	0.653	0.633	0.617
0,01		0.993	0.942	0.883	0.833	0.793	0.761	0.734	0.711	0.691
0,05	4	0.906	0.768	0.684	0.629	0.589	0.560	0.536	0.518	0.502
0,01		0.968	0.864	0.781	0.721	0.676	0.641	0.613	0.590	0.570
0,05	5	0.841	0.684	0.598	0.544	0.506	0.478	0.456	0.439	0.424
0,01		0.928	0.789	0.696	0.633	0.588	0.553	0.526	0.504	0.485
0,05	6	0.781	0.616	0.532	0.480	0.445	0.418	0.398	0.382	0.368
0,01		0.883	0.722	0.626	0.563	0.520	0.487	0.461	0.440	0.423
0,05	7	0.727	0.561	0.480	0.431	0.397	0.373	0.354	0.338	0.326
0,01		0.838	0.664	0.568	0.508	0.466	0.435	0.411	0.391	0.375
0,05	8	0.680	0.516	0.438	0.391	0.359	0.336	0.318	0.304	0.293
0,01		0.794	0.615	0.521	0.463	0.423	0.393	0.370	0.352	0.337
0,05	9	0.638	0.477	0.403	0.358	0.328	0.307	0.290	0.277	0.266
0,01		0.754	0.573	0.481	0.425	0.387	0.359	0.338	0.321	0.307
0,05	10	0.602	0.445	0.373	0.331	0.303	0.282	0.267	0.254	0.244
0,01		0.717	0.536	0.447	0.393	0.357	0.331	0.311	0.295	0.281