

**TRABAJO DE FIN DE GRADO**

**GRADO EN CIENCIA Y TECNOLOGÍA DE LOS ALIMENTOS**



**UNIVERSIDAD DE BURGOS**

**ESTUDIO DEL EXTRACTO ACUOSO DE PIMENTONES CON TÉCNICAS  
EXPLORATORIAS Y ANÁLISIS DE SU POSIBLE ADULTERACIÓN CON  
COLORANTES ARTIFICIALES**

Tutoras: Ana Herrero Gutiérrez y Celia Reguera Alonso

**PATRICIA SÁNCHEZ DE BENITO**

**Febrero 2019**

# Índice

Resumen.....	1
Abstract.....	1
1. Introducción.....	2
2. Objectives.....	5
3. Teoría.....	5
3.1 Técnicas exploratorias.....	5
3.1.1 Cluster análisis.....	5
3.1.2 Análisis de componentes principales.....	7
3.2 Regresión por mínimos cuadrados parciales (PLS).....	8
4. Experimental.....	9
4.1 Muestras y reactivos.....	9
4.2 Procedimiento experimental.....	10
5. Resultados y discusión.....	10
5.1 Análisis exploratorio.....	10
5.1.1 Señales espectrales.....	10
5.1.2 Variables CIELab.....	14
5.2 Análisis de regresión.....	16
5.2.1 Estudio inicial.....	16
5.2.2 Modelos de calibrado a partir de un único pimentón.....	18
5.2.3 Modelos de calibrado conjuntos.....	20
6. Conclusiones.....	22
7. Bibliografía.....	23

## Resumen

El pimentón es un producto alimentario susceptible de ser adulterado, por ejemplo, con la adición de colorantes artificiales con el objetivo de avivar su color y asemejarlo más al producto auténtico. En este trabajo, se ha abordado la adulteración del pimentón con dos colorantes alimentarios, la tartrazina (E-102) y el amarillo anaranjado (E-110), mediante el análisis del extracto acuoso del pimentón.

En una primera etapa se lleva a cabo un análisis exploratorio, basado en el análisis de agrupamientos y en el análisis de componentes principales, para estudiar si las muestras de pimentón se pueden agrupar según su origen o tipo, utilizando para ello los espectros de absorción molecular en el visible de sus extractos acuosos. Asimismo, este análisis también se lleva a cabo con las variables CIELab.

Finalmente, se aborda la determinación de los dos colorantes alimentarios, en extractos acuosos de pimentones adulterados, mediante regresión por mínimos cuadrados parciales (PLS) con el objetivo de detectar adulteraciones de hasta un 3,2%.

## Abstract

Paprika is a food product that is susceptible to being adulterated, e.g. by adding artificial colours with the aim of enhancing its colour and resembling the appearance of the genuine product. The adulteration of paprika with two food colours, tartrazine (E-102) and sunset yellow (E-110) by the analysis of the aqueous extract is approached in this work.

In a first step, an exploratory analysis, based on cluster analysis and principal component analysis, is carried out to study if paprika samples may be grouped according to their origin or type, using the visible molecular absorption spectra of the aqueous extract. Likewise, the analysis is also performed with the CIELab variables.

Lastly, the determination of the two food colours is conducted in the aqueous extracts of adulterated paprika sample by means of partial least squares regression with the aim of detecting adulterations up to 3,2%.

# 1. Introducción

El pimiento (*Capsicum annum*), taxonómicamente pertenece a la familia Solanáceas del orden *Tubiflora*. Se trata de una planta herbácea, con sistema radicular profundo, provisto de un gran número de raíces. El fruto es una baya semicartilaginosa, sin jugo interno, de color rojo o en ciertos casos amarillo. Las semillas son redondeadas, de color amarillo pálido [1]. El origen geográfico del pimiento puede situarse en Centro-Sudamérica, la planta llegó a Europa tras el descubrimiento del continente americano.

Entre los compuestos con marcadas propiedades biológicas del pimentón caben destacar los relacionados con su poder colorante y con su sabor. En el primer caso se incluyen los carotenoides [2, 3] entre los que destacan: xantofilas,  $\beta$ -caroteno, capsantina, capsorrubina, criptoxantina, zeaxantina, anteraxantina, violanxantina, noxantina; en el segundo grupo se encuentran los capsaicinoides (capsaicina, dihidrocapsaicina, entre otros).

En cuanto a sus propiedades picantes, los pimentones se pueden clasificar en:

- Variedades muy picantes, con un contenido en capsaicinoides entre el 0,3 y el 1 % en peso seco.
- Variedades picantes, con un contenido en capsaicinoides entre 0,01 y 0,3% en peso seco del pimiento.
- Variedades dulces o moderadamente picantes, con un contenido en capsaicinoides inferior al 0,01% en peso seco del pimiento. Estos son de interés tanto por sus propiedades colorantes como por su sabor.

Uno de los tipos de consumo del pimiento es como pimentón, que se utiliza ampliamente como colorante natural para modificar el color de los alimentos, mejorar su aspecto y, en algunos casos, conferir características organolépticas particulares relacionadas con el sabor y aroma.

El proceso de obtención del pimentón consta principalmente de dos operaciones, una primera deshidratación o secado de los frutos maduros y una posterior molturación (Figura 1). Para desecar el pimiento se utilizan diversos métodos como secado al sol, directo o bajo cubierta de plástico, secado por combustión de leña o al humo y secado mediante aire caliente [4].



Figura 1. Esquema de la producción del pimentón.

La importancia económica del pimentón es muy elevada ya que es una de las principales fuentes de riqueza para muchas regiones tanto del mundo como de España donde hay varias Denominaciones de Origen Protegidas (D.O.P):

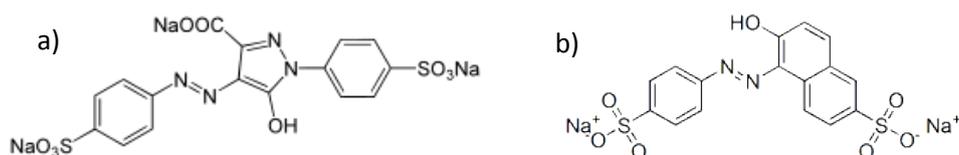
-D.O.P. "Pimentón de la Vera" [5]. El Pimentón de la Vera es el producto obtenido de la molienda de frutos totalmente rojos, de las variedades del grupo de los "Ocaleas", jaranda, jariza, jeromín, y de la variedad bola, pertenecientes a las especies *Capsicum annum L.* y *Capsicum longum L.*, recolectados maduros, sanos, limpios, con el color característico de la variedad, libres de ataques de plagas o enfermedades, secados con leña de encina y/o roble, por el sistema tradicional de la Vera, y que proceda de dicha zona de producción específica.

-D.O.P. "Pimentón de Murcia" [6]. En la comarca del campo de Cartagena, el clima y el suelo son muy apropiados para el cultivo de pimientos de la especie *Capsicum annum L.*, de la variedad bola. Recolectados maduros, sanos, limpios y secos, libres de ataques de plagas o enfermedades, se muelen y se obtiene un pimentón dulce, de sabor fuerte, penetrante y característico, de gran poder colorante, graso y sabroso. El color rojo brillante es muy estable.

La normativa de referencia aceptada en el control de calidad del pimentón es la especificación técnica UNE-EN ISO 7540:2010 [7], en la cual se muestran las especificaciones que debe cumplir.

La detección de posibles fraudes y adulteraciones en el pimentón es de vital importancia para los consumidores. El objetivo de la adulteración mediante la adición de colorantes es imitar el producto auténtico, en este caso el pimentón.

Los colorantes son una clase de aditivos alimentarios sin valor nutricional, que se agregan a los alimentos con el objetivo de proporcionar color, aumentando así la aceptación por parte del consumidor. Entre los colorantes más utilizados en la industria alimentaria se encuentran la tartrazina o E-102 (Figura 2a) y el amarillo anaranjado o E-110 (Figura 2b), ambos contienen un grupo azo (-N=N-) [8].



**Figura 2.** Estructura química de a) tartrazina (E-102) y b) amarillo anaranjado (E-110).

El E-110, es un colorante azoico sintético y soluble en agua. Sin embargo, presenta un problema ya que, mezclado con otros aditivos, puede presentar reacciones de sensibilidad como enrojecimiento facial, picazón o hiperactividad en niños.

El E-102, es un colorante azoico producido por acoplamiento azoico de derivados del petróleo. Es soluble en agua, estable al calor, luz y ácidos. Puede provocar reacciones alérgicas, asma e hipersensibilidad al ácido salicílico [9].

La utilización de colorantes está legislada por el Reglamento (CE) nº 1333/2008 sobre aditivos alimentarios [10], el cual incluye las normas de utilización para asegurar la salud de los consumidores. Este reglamento establece listas comunitarias de los diferentes aditivos alimentarios autorizados, que figuran en los anexos I y II, las normas para su etiquetado, así como las condiciones de utilización. En el año 2011, la lista de colorantes autorizados fue modificada en el Reglamento (UE) nº 1129/2011 de la Comisión, de 11 de noviembre de 2011 [11].

Existen diversos métodos para el análisis del pimentón, siendo los principales los relacionados con la medida del color a través de parámetros como el índice ASTA [12], que da idea del contenido total de carotenoides de un pimentón [13], y de las variables CIELab (Commission internationale de l'éclairage) [14] que forman un espacio tridimensional para describir todos los colores que describe el ojo humano.

El espacio CIELab tridimensional (Figura 3) está definido por tres vectores que son la claridad,  $L^*$  ( $L^*= 0$  negro,  $L^*=100$  blanco), la posición entre el verde y el rojo o cromaticidad rojo/verde, definida por  $a^*$  ( $a^*<0$  verde,  $a^*>0$  rojo) y  $b^*$ , cromaticidad amarillo/azul que hace referencia a la posición entre el amarillo y el azul ( $b^*< 0$  azul,  $b^*>0$  amarillo) [15].

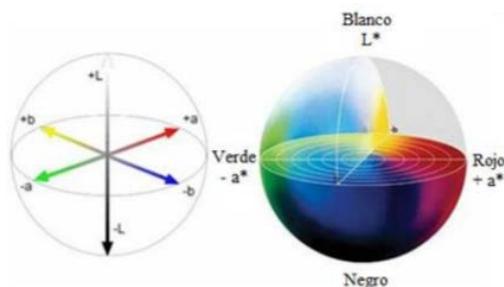


Figura 3. Espacio CIELab.

Las medidas espectrales de absorción molecular UV-Vis junto con técnicas de reconocimiento de pautas han sido muy utilizadas en la caracterización y clasificación de alimentos como el vino y otras bebidas alcohólicas [16, 17], café [18], té [19] y aceite de oliva [20].

En el caso del pimentón, se han realizado algunos estudios para la identificación del origen geográfico y la autenticidad del producto basados en medidas de espectroscopia atómica [21, 22] y en modelos basados en coordenadas ASTA y CIELab [23].

Este trabajo se presenta como continuación de un trabajo previo [24] en el que se abordó la extracción de colorantes artificiales adulterantes del pimentón utilizando como medio extractante solventes orgánicos, y se concluyó que además de la extracción de los colorantes artificiales también se extraían otros compuestos coloreados del pimentón que interferían en el análisis. En el presente trabajo, puesto que los adulterantes considerados (E-110 y E-102) son solubles en agua, se lleva a cabo la extracción de los colorantes artificiales en medio acuoso ya que los carotenoides del pimentón no son hidrosolubles.

## 2. Objectives

The objectives of this work have been:

- Use the principal component analysis and cluster analysis with the CIELab variables to group paprika samples.
- Study and evaluate the presence of adulterants such as E-102 and E-110 in aqueous extract of paprika.
- Validate the multivariate analytical methods developed and detect the presence of outliers.

## 3. Teoría

### 3.1 Técnicas exploratorias

#### 3.1.1 Cluster análisis

El análisis "Cluster" o de agrupamientos, constituye una herramienta de descripción de datos que permite la clasificación de muestras u objetos en distintos grupos, de manera que exista la mayor homogeneidad posible dentro de los grupos, con respecto a ciertas variables, variables predictoras, que pueden ser cuantitativas o cualitativas.

La clasificación de objetos se realiza a partir de las variables observadas en ellos, por lo que es condición primordial realizar una buena elección de las variables iniciales, así como también elegir una medida de homogeneidad o similitud adecuada para la situación que se esté analizando. No existe una única medida de similitud, ni tampoco es único el método de agrupar observaciones en distintos clusters.

#### Distancias y similitudes

La disimilitud es una medida de lo distintos que son dos objetos. Suponiendo los objetos  $x_k$  y  $x_h$ , definidos por una serie de variables continuas, de manera que  $x_k = (x_{k1}, x_{k2}, \dots, x_{kp})$  y  $x_h = (x_{h1}, x_{h2}, \dots, x_{hp})$ . Si las variables son continuas, la disimilitud más usada es la distancia ( $d$ ) entre objetos; las distancias más empleadas en análisis de agrupamientos son:

- Distancia euclídea

$$d(x_k, x_h) = \sqrt{(x_{k1} - x_{h1})^2 + (x_{k2} - x_{h2})^2 + \dots + (x_{kp} - x_{hp})^2}$$

- Distancia euclídea al cuadrado

$$d(x_k, x_h)^2 = \sum_{j=1}^p (x_{kj} - x_{hj})^2$$

La transformación de las distancias (disimilitudes) en similitudes ( $s$ ) queda como:

$$s_{kh} = s(x_k, x_h) = 1 - \frac{d(x_k, x_h)}{d_{\max}}$$

lo que da lugar a la matriz de similaridades  $S = (s_{kh})$  que es una matriz cuadrada, en general simétrica, y con  $d_{\max}$  ó 1 en la diagonal principal.

### Métodos jerárquicos

Estos métodos tienen por objetivo agrupar clusters para formar uno nuevo o bien separar alguno ya existente para dar origen a otros dos, de tal forma que se minimice alguna función distancia o bien se maximice alguna medida de similaridad. Los métodos jerárquicos se subdividen a su vez en aglomerativos y disociativos [25].

Los aglomerativos comienzan el análisis con tantos grupos como objetos haya en el estudio. A partir de ahí se van formando grupos de forma ascendente hasta que, al final del proceso, todos los casos están englobados en un mismo conglomerado.

Los métodos disociativos realizan el proceso inverso al anterior. Empiezan con un conglomerado que engloba a todos los objetos. A partir de este grupo inicial se van formando, a través de sucesivas divisiones, grupos cada vez más pequeños. Al final del proceso se tienen tantos grupos como objetos hay en el estudio.

Centrándonos en los métodos aglomerativos y en un conjunto de  $n$  objetos, el primer nivel de agrupamiento,  $K= 0$ , estará formado por  $n$  grupos. En el siguiente nivel se agruparán aquellos dos objetos que tengan mayor similaridad (o menor distancia), resultando así  $n-1$  grupos; a continuación, y siguiendo con la misma estrategia, se agruparán en el nivel posterior, aquellos dos objetos (o clusters ya formados) con menor distancia o mayor similaridad; de esta forma, en el nivel  $L$  tendremos  $n-L$  grupos formados. Si se continúa agrupando de esta forma, se llega al nivel  $L= n-1$  en el que solo hay un grupo formado por todos los objetos.

Esta manera de formar nuevos grupos tiene la particularidad de que, si en un determinado nivel se agrupan dos clusters, estos quedan ya jerárquicamente agrupados para el resto de los niveles.

Los métodos jerárquicos permiten la construcción de un árbol de clasificación, que recibe el nombre de dendrograma (Figura 4), en el cual se puede seguir de forma gráfica el procedimiento de unión seguido, mostrando qué grupos se van uniendo, en qué nivel concreto lo hacen, así como el valor de la medida de asociación entre los grupos cuando estos se agrupan.

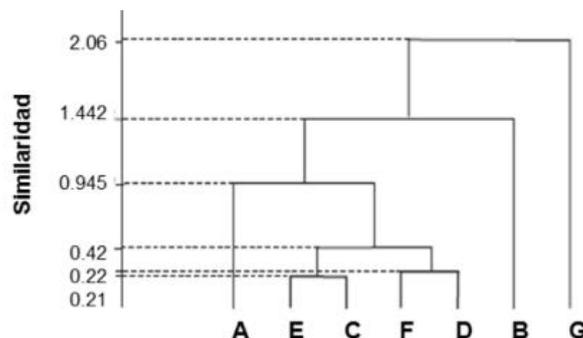


Figura 4. Dendrograma.

### Criterios de agrupamiento

Hay diversos criterios para ir formando los conglomerados en un procedimiento jerárquico; todos estos criterios se basan en la matriz de distancias o similaridades, a continuación se explican los utilizados en el análisis exploratorio.

1. Método del vecino más próximo (simple linkage): en este método se considera que la distancia o similaridad entre dos clusters viene dada, respectivamente, por la mínima distancia (o máxima similaridad) entre sus componentes. Una vez calculadas las distancias entre todos los objetos, o clusters ya formados, se unirán aquellos cuya distancia sea mínima.

2. Método del vecino más lejano (complete linkage): en este método se considera que la distancia entre dos clusters hay que medirla atendiendo a sus elementos más dispares; o sea, la distancia entre clusters viene dada por la máxima distancia (o mínima similaridad) entre sus componentes. En este caso, una vez calculadas las distancias entre todos los objetos, o clusters ya formados, se unirán aquellos cuya distancia sea mínima.

3. Método de Ward: es un procedimiento jerárquico en el cual, en cada etapa, se unen los dos clusters para los cuales se tiene el menor incremento en el valor total de la suma de cuadrados intra-cluster (SSI), que se evalúa a través de las diferencias, dentro de cada cluster, de cada individuo al centroide del cluster.

$$SSI = \sum_{k=1}^K \sum_{i \in C_k} \sum_{j=1}^p (x_{ij} - \bar{x}_j(k))^2$$

El proceso comienza con n clusters, cada uno de los cuales está compuesto por un solo individuo, por lo que cada individuo coincide con el centroide del cluster. Se puede demostrar que el menor incremento de los errores cuadráticos es proporcional a la distancia euclídea al cuadrado de los centroides de los clusters unidos

### 3.1.2 Análisis de componentes principales

Cuando se tienen diversas variables en cada muestra, estas últimas se encuentran distribuidas en el espacio formando una nube de tantas dimensiones como variables predictoras se estén considerando. El análisis de componentes principales es una técnica de análisis de datos que facilita la búsqueda de posibles pautas o tendencias en un conjunto de datos multivariante, al reducir la dimensionalidad de esa nube de puntos y establecer relaciones entre objetos y variables para extraer su estructura propia [26].

La técnica consiste en crear un nuevo sistema de coordenadas en el que se sustituirán las p variables predictoras ( $X_1, X_2, \dots$ ), por las puntuaciones en las q componentes principales, que son combinaciones lineales de las variables originales:

$$CP_q = I_{q1}X_1 + I_{q2}X_2 + \dots + I_{qp}X_p$$

Los coeficientes o "loadings" ( $I_{ij}$ ) indican los pesos de cada variable dentro de la combinación lineal de cada componente principal. Para evitar que las componentes principales se vean influenciadas por las diferentes escalas de las variables, es habitual estandarizar o autoescalar las variables, de modo que su varianza sea 1 y se encuentren centradas en el origen.

La primera componente principal sigue la dirección en la que la variabilidad de los puntos que forman la nube es máxima, por lo que se buscan los coeficientes de la combinación lineal correspondiente a la dirección de máxima elongación de la nube de puntos.

La segunda componente principal se construye de modo que sea ortogonal a la primera y en la siguiente dirección de máxima varianza, es decir, en la dirección no correlacionada con la anterior que explica el mayor porcentaje de la varianza no explicada por la primera componente. De este modo se construyen las restantes componentes principales. El número de componentes principales que se pueden construir está limitado bien por el número de variables predictoras o bien por el número de objetos.

$$CP_1 = l_{11}X_1 + l_{12}X_2 + \dots + l_{1p}X_p \text{ tal que } \text{Var}(CP_1) = \lambda_1 \text{ sea máxima}$$

$$CP_2 = l_{21}X_1 + l_{22}X_2 + \dots + l_{2p}X_p \text{ tal que } \text{Var}(CP_2) = \lambda_2 \text{ sea máxima y con la restricción } \text{Corr}(CP_1, CP_2) = 0$$

...

$$CP_q = l_{q1}X_1 + l_{q2}X_2 + \dots + l_{qp}X_p \text{ tal que } \text{Var}(CP_q) = \lambda_q \text{ sea máxima y con la restricción } \text{Corr}(CP_q, CP_1) = \text{Corr}(CP_q, CP_2) = \dots = \text{Corr}(CP_q, CP_{q-1}) = 0.$$

Si es posible describir con suficiente precisión los valores de las  $p$  variables por un pequeño subconjunto ( $k < p$ ) de ellas, se consigue reducir la dimensión del problema, aunque suponga la pérdida de una pequeña parte de la información. Se asume que esa parte de la información no va a ser relevante o que estará relacionada con ruido. De este modo, es posible representar la información en un espacio dimensional menor, lo que facilita la interpretación de las componentes principales y su relación con las variables predictoras y el conjunto de datos en estudio.

### 3.2 Regresión por mínimos cuadrados parciales (PLS)

Cuando se desean determinar varias sustancias que absorben en una misma región espectral, es decir, con señales espectrales no específicas, el problema no se puede abordar mediante un análisis univariante. Una alternativa es realizar una calibración multivariante en la que se considere el espectro completo.

Si se trata de señales espectrales no es posible llevar a cabo una regresión lineal múltiple de la concentración del analito de interés (variable respuesta) sobre las absorbancias registradas (variables predictoras), ya que las altas correlaciones existentes entre este tipo de variables y las posibles colinealidades originan modelos inestables y con poca capacidad de predicción.

Los modelos de regresión sobre variables latentes permiten llevar a cabo determinaciones aun cuando las variables predictoras, no sean completamente específicas o estén altamente correlacionadas. Entre estas técnicas se encuentran la regresión sobre componentes principales (principal component regression, PCR) y la regresión por mínimos cuadrados parciales (partial least squares regression, PLS). Ambas construyen modelos de regresión utilizando la información contenida en todo el espectro para obtener unas nuevas variables, las variables latentes, buscando tener una alta capacidad predictiva.

La regresión sobre las componentes principales consiste en hacer una regresión lineal múltiple con las puntuaciones de las  $k$  componentes principales, de este modo se obvian los problemas derivados de la correlación y la colinealidad antes mencionados y, además, se reduce la inclusión de ruido en el modelo de calibrado. Sin embargo, en este caso, al no existir ninguna relación entre el modo en que se obtienen las componentes principales y la variable respuesta se puede producir un subajuste.

Como consecuencia de esto se han desarrollado modelos de regresión PLS que construyen variables latentes buscando direcciones de máxima variabilidad en el espacio de las variables predictoras, pero evitando las no correlacionadas con la respuesta [27]. Al igual que en el caso anterior, las variables latentes se obtienen mediante combinaciones lineales de las variables originales, los coeficientes de dichas combinaciones se denominan pesos, las coordenadas de cada muestra en el nuevo espacio vectorial son las puntuaciones.

La selección del número de variables latentes a considerar en el modelo es un punto crítico. Se puede llevar a cabo mediante validación cruzada, que se basa en predecir la respuesta en alguna/s muestra/s con un modelo PLS construido con las muestras restantes. Esto se repite de forma que se predice la respuesta para todas las muestras. Las mejores predicciones sobre estas muestras indicarán el número de variables latentes óptimo. Para conjuntos de datos reducidos, únicamente se deja fuera una muestra y la respuesta se predice con las demás, el método se denomina leave-one-out. Si el número de muestras es más elevado, se puede utilizar el método de persianas venecianas, en el que las muestras a predecir se seleccionan en bloques, con una alternancia determinada (cada muestra test se determina seleccionando cada objeto k-ésimo del conjunto de datos, comenzando por las muestras de 1 a k).

Por otro lado, para el diagnóstico de datos anómalos se utilizan dos índices: los estadísticos Q y  $T^2$  de Hotelling. La suma de cuadrados residual (estadístico Q) indica el residuo ortogonal de cada objeto a las variables latentes consideradas en el modelo, mientras que el estadístico  $T^2$  de Hotelling denota la distancia al centroide de ese espacio multidimensional. Un objeto se considera anómalo si, para un nivel de confianza del 95%, los valores de ambos índices superan simultáneamente los valores críticos correspondientes.

## 4. Experimental

### 4.1 Muestras y reactivos

Se han analizado en total 40 muestras diferentes de pimentón (Figura 5), que se pueden clasificar según el tipo de pimentón en agridulce (A), dulce (D) o picante (P); y según el origen de procedencia en “de la Vera” (V), diferente al “de la Vera” (N) o si se desconocido (D) si se desconoce su origen. De los 40 pimentones analizados, 27 eran pimentones dulces, 12 eran picantes y uno agridulce. En cuanto a su origen, 11 pertenecían a la D.O.P “de la Vera”, 21 no pertenecían a dicha denominación y de los 8 restantes no se conocía su origen.



Figura 5. Algunas de las muestras de pimentón analizadas.

Para estudiar la adulteración del pimentón con colorantes artificiales alimentarios se utilizó tartrazina E-102, de la marca Aldrich con una pureza del 90% y amarillo anaranjado E-110, de la marca Fluka con una pureza del 100%.

Las medidas espectrofotométricas se han realizado con un espectrofotómetro UV-visible Cary 50 Conc de Varian, tomando valores de absorbancia cada 5 nm entre las longitudes de onda de 380 y 800 nm.

El tratamiento de datos se llevó a cabo utilizando Statgraphics [28] y Matlab [29]. Con Statgraphics se realizó el análisis de componentes principales (PCA), y el análisis cluster, y con Matlab la regresión por mínimos cuadrados parciales (PLS).

## **4.2 Procedimiento experimental**

Se prepararon las disoluciones acuosas de los colorantes en estudio (E-102 y E-110), con una concentración de  $25 \text{ mg L}^{-1}$ , posteriormente se diluyó la hasta obtener una concentración de  $10 \text{ mg L}^{-1}$ .

De cada uno de los pimentones se pesaron 0,03 gramos y se adicionaron 20 mL de agua destilada, seguidamente se mantuvo en agitación en un baño de ultrasonidos P-selecta durante 10 minutos. Pasado este tiempo, la muestra se filtró a través de papel de filtro. El extracto acuoso obtenido se filtró de nuevo, antes de obtener la medida espectrofotométrica, utilizando un filtro de nylon de  $0,45 \mu\text{m}$  que permite eliminar las posibles partículas en suspensión de menor tamaño.

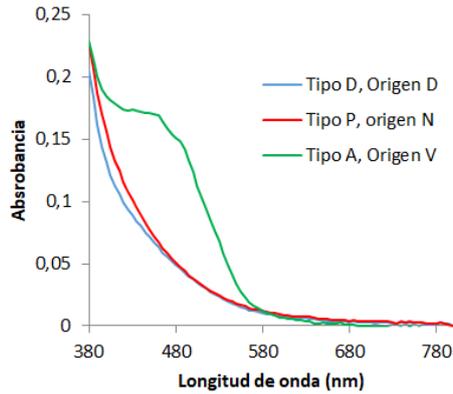
## **5. Resultados y discusión**

### **5.1 Análisis exploratorio**

#### **5.1.1 Señales espectrales**

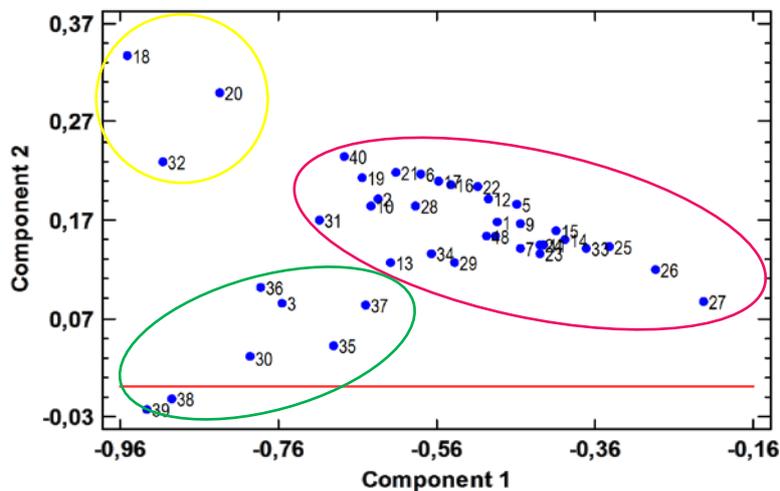
Con fin de comprobar si las muestras de pimentón se pueden agrupar según el contenido de su extracto acuoso se registraron los espectros de dichos extractos y se llevó a cabo un análisis exploratorio. En la Figura 6 se pueden observar algunos de los espectros registrados para cada tipo de pimentón según su origen y tipo.

A continuación, se realizó el análisis de componentes principales con las absorbancias originales, sin ningún tipo de pretratamiento. El conjunto de datos constaba de una matriz de 40 objetos (extractos procedentes de los diferentes pimentones) y 84 variables (absorbancias).

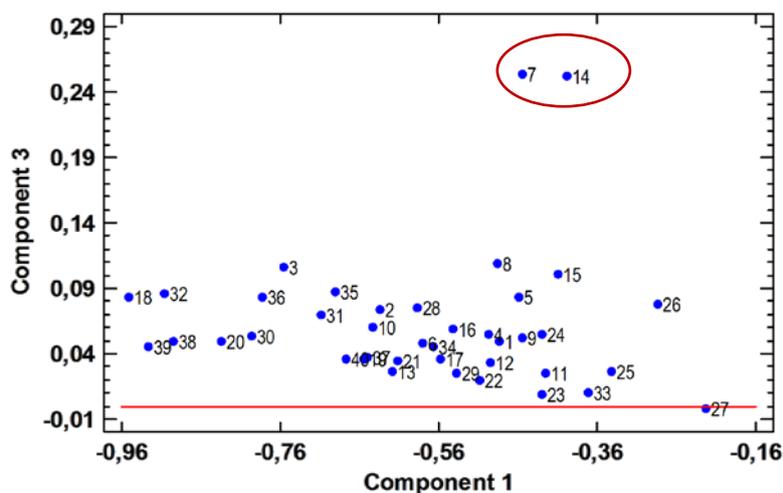


**Figura 6.** Espectros según el tipo y origen de pimentón. Azul: dulce y de origen desconocido; rojo: picante y no de la Vera; y verde: agrisulce y de la Vera.

Las tres primeras componentes principales explican un 99,53%, si bien, con las dos primeras componentes principales ya se explica un 92,68% de varianza. La representación de las puntuaciones de las muestras en el plano formado por las dos primeras componentes principales permite diferenciar tres grupos, como se observa en la Figura 7. Respecto a la representación en el plano formado por las componentes 1 y 3, se puede observar en la Figura 8 que la tercera componente principal está relacionada con dos pimentones, las muestras 7 y 14.

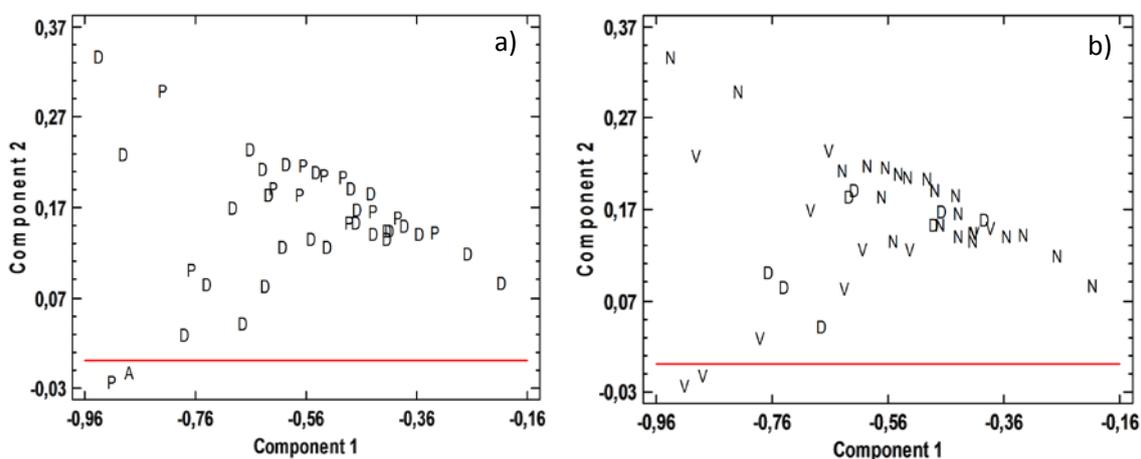


**Figura 7.** Puntuación de las muestras de pimentón en las dos primeras componentes principales.



**Figura 8.** Puntuaciones de las muestras de pimentón en las componentes principales 1 y 3.

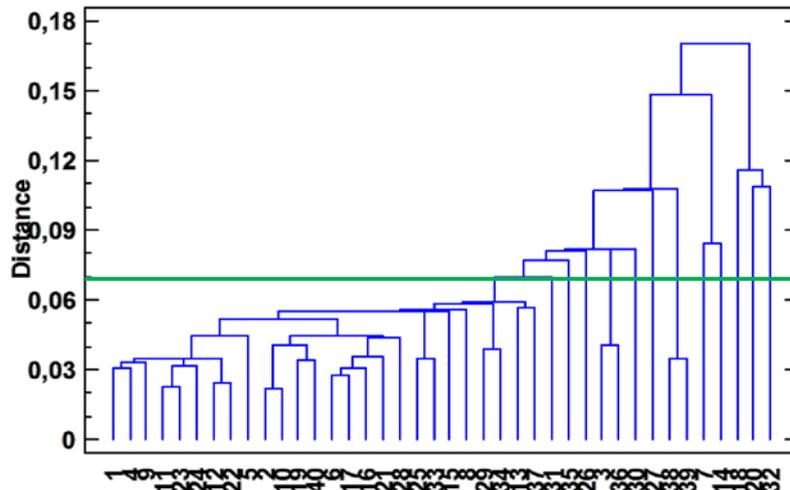
Comprobando el tipo y el origen de los pimentones incluidos en cada grupo se puede establecer que no existe relación entre los grupos formados y el tipo de pimentón, como se muestra en la Figura 9a. Sin embargo, en cuanto al origen, se observa que las muestras que no son “de la Vera” (N) se incluyen en dos de los grupos, como se muestra en la figura 9b y se sitúan en la diagonal del gráfico, mientras que el tercer grupo está formado solo por muestras “de la Vera” (V) y de origen desconocido (D) que podrían pertenecer también a esta D.O.P. En cuanto a los pimentones 7 y 14, relacionados con la tercera componente principal, ambos son dulces, pero de origen distinto.



**Figura 9.** Puntuación de las dos primeras componentes principales respecto (a) el tipo y (b) origen del pimentón.

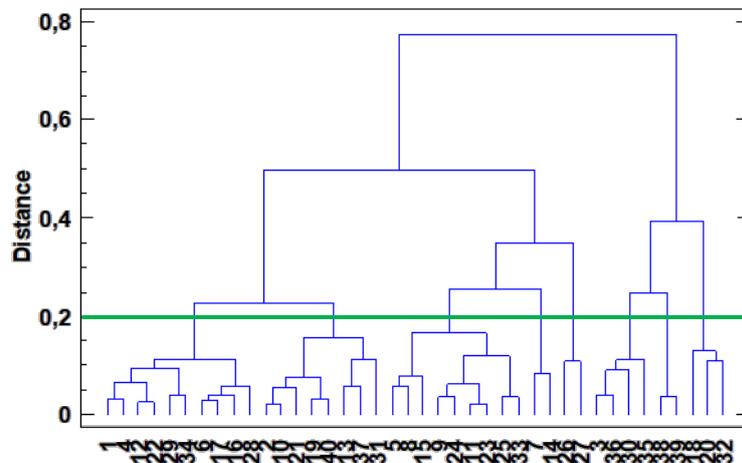
A continuación, se llevó a cabo un análisis cluster con las 40 muestras y las 84 variables originales. La medida de la similitud empleada fue la distancia euclídea y se estudiaron distintos métodos de agrupamiento como el del vecino próximo, el del vecino más lejano y el método de Ward.

El método del vecino más próximo muestra muchos grupos formados, cada uno de ellos, por un número pequeño de muestras. Con una distancia de 0,07 se forman 12 grupos diferentes, como se observa en la Figura 10.



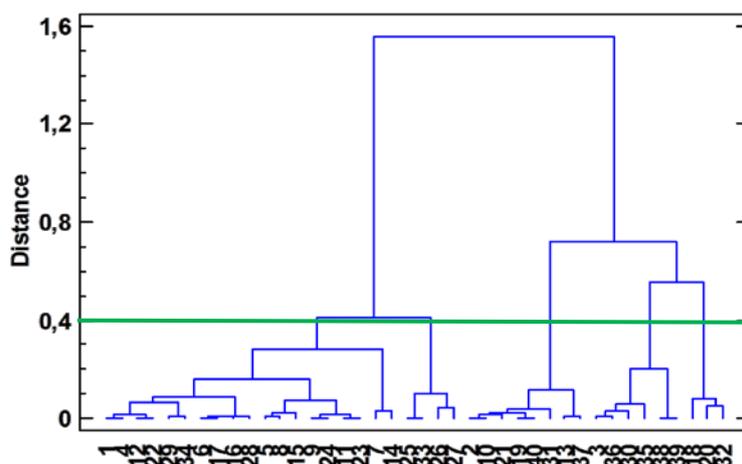
**Figura 10.** Dendrograma del método del vecino más próximo para la agrupación de los pimentones.

Con el método del vecino más lejano la distancia entre clusters viene dada por la máxima distancia (o mínima similitud) entre sus componentes, por esta razón se obtienen menos grupos que en el caso anterior. En la Figura 11 se muestra que con una distancia de 0,2 se forman 8 grupos.



**Figura 11.** Dendrograma del método del vecino más lejano para la agrupación de los pimentones.

Puesto que lo que se quiere es encontrar un número no muy alto de agrupamientos para poder seleccionar pimentones representativos de ellos, el método de Ward permitiría, para valores de distancia relativamente bajos (alta similitud), obtener un número de agrupamientos bajo. Concretamente para una distancia de 0,4 los pimentones quedarían clasificados en 5 grupos, como se muestra en Figura 12.



**Figura 12.** Dendrograma del método de Ward para la agrupación de los pimentones.

El análisis cluster llevado a cabo con los espectros de los extractos, al igual que el análisis de las componentes principales, tampoco permite diferenciar entre los tipos de pimentones analizados, es decir, no conduce a agrupaciones relacionadas con el tipo de pimentón. Además, en este caso tampoco se pueden establecer grupos relacionados con distintos orígenes de los pimentones.

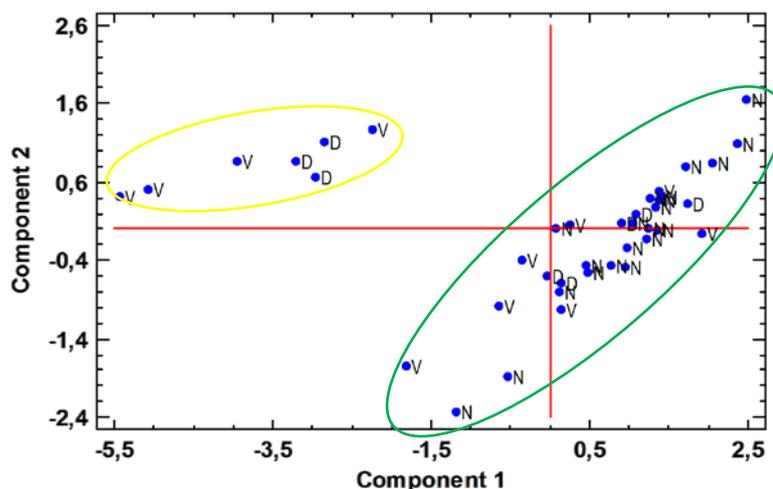
### 5.1.2. Variables CIELab

En una segunda etapa del estudio, se aplican las técnicas de análisis exploratorio a la matriz de datos formada por las 40 muestras de pimentón y 5 variables CIELab correspondientes a la claridad ( $L^*$ ), cromaticidad rojo/verde ( $a^*$ ), cromaticidad amarillo/azul ( $b^*$ ), croma ( $C^*$ ) y tono angular ( $h^*$ ), todas ellas estimadas a partir de las señales espectrales. Realizando el análisis componentes principales, con las variables estandarizadas, se tiene que con las cinco primeras componentes principales se explica el 100% de varianza, como se muestra en la Tabla 1.

En el plano formado por las dos primeras componentes principales (Figura 13), se observan dos grupos diferenciados por las puntuaciones que tienen las muestras en la primera componente principal. Así, el grupo con mayores puntuaciones en esa componente incluye muestras de diversos orígenes, pero todas las que no son “de la Vera” están incluidas en él. Mientras que en el grupo con menores puntuaciones en la primera componente principal sólo se incluyen muestras “de la Vera” y de origen desconocido.

**Tabla 1.** Análisis de componentes principales con las variables CIELab.

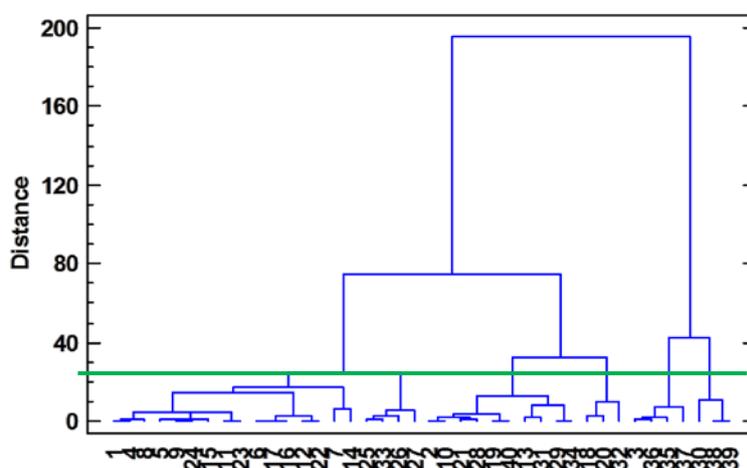
Componente principal	Autovalor	Varianza explicada (%)	Varianza acumulada (%)
1	3,95	79,08	79,08
2	0,72	14,39	93,47
3	0,23	4,61	98,08
4	0,09	1,92	99,99
5	$1,06 \cdot 10^{-4}$	$2 \cdot 10^{-3}$	100



**Figura 13.** Puntuaciones de las muestras en las dos primeras componentes principales.

Llevando a cabo el análisis cluster con esta nueva matriz de datos y utilizando el método de Ward como método jerárquico de agrupamiento, se tiene que para una distancia de 25 se establecen 6 grupos diferentes, los cuales se aprecian en la Figura 14. Aunque las variables CIELab describen los colores que el ojo humano es capaz de apreciar, en este caso no se observa, a simple vista una coloración parecida entre los pimentones de un mismo grupo (Figura 15). Esto puede ser debido a que los compuestos responsables del color en los pimentones son poco solubles en medio acuoso y, por tanto, las variables CIELab utilizadas en este estudio estarán relacionadas con otros compuestos coloreados del pimentón presentes en el extracto acuoso.

Sin embargo, sí que se puede afirmar que los agrupamientos establecidos son muy similares a los obtenidos en el análisis cluster realizado con los datos espectrales originales. En la Tabla 2 se muestra una comparación de las agrupaciones obtenidas con el método de Ward aplicado a las absorbancias originales y a las variables CIELab para las 40 muestras de pimentón.



**Figura 14.** Dendrograma del método de Ward con las variables CIELab.



**Figura 15.** Pimentones del mismo grupo según el método de Ward utilizando las variables CIELab.

**Tabla 2.** Grupos formados según el método de Ward con las absorbancias originales y con las variables CIELab.

Absorbancias originales						
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	
<b>Muestras</b>	1, 4, 12, 22, 29, 34, 6, 17, 16, 28, 5, 8, 15, 9, 24, 11, 23, 7, 14	25, 33, 26, 27	2, 10, 21, 19, 40, 31, 13, 37	3, 36, 30, 35, 38, 39	18, 20, 32	
<b>Tipo</b>	D y P	D y P	D y P	A, D y P	D y P	
<b>Origen</b>	D, N y V	N	D, N y V	D y V	N y V	

Variables CIELab						
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6
<b>Muestras</b>	1, 4, 8, 5, 9, 24, 15, 11, 23, 6, 17, 16, 12, 22, 7, 14	25, 33, 26, 27	2, 10, 21, 28, 19, 40, 13, 31, 29, 34	18, 20, 32	3, 36, 35, 37	30, 38, 39
<b>Tipo</b>	D y P	D y P	D y P	D y P	D y P	D, A y P
<b>Origen</b>	D, N y V	N	D, N y V	N y V	D y V	V

## 5.2 Análisis de regresión

En la última parte del trabajo se aborda el estudio de la adulteración de pimentón con los colorantes E-102 y E-110. Para llevar a cabo el análisis de regresión se ha escogido un pimentón de los diferentes grupos formados a través del análisis de agrupamientos con las variables CIELab. Los pimentones elegidos fueron los números: 7, 30, 25, 18, 19, 1.

### 5.2.1 Estudio inicial

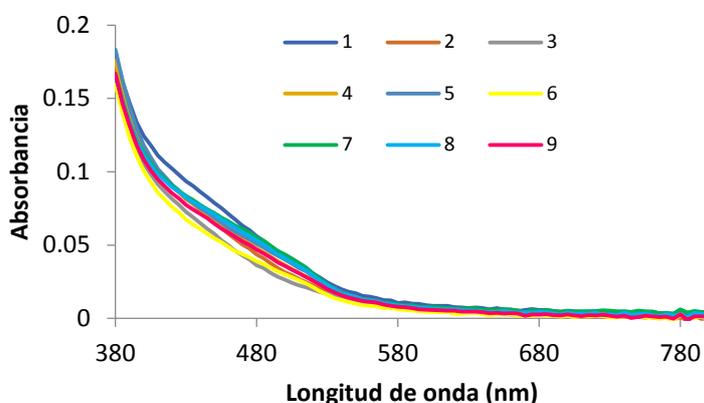
Al llevar a cabo el análisis de adulteración cabe esperar que la matriz de pimentón ejerza un efecto interferente sobre la señal analítica considerada. Por ello, se plantea realizar la calibración utilizando muestras de extractos de pimentones enriquecidas con concentraciones conocidas de los colorantes en estudio para así poder modelar, mediante la regresión PLS, el efecto matriz.

Como conjunto de calibrado se consideran extractos acuosos de 6 pimentones distintos enriquecidos con concentraciones variables de cada uno de los colorantes estudiados siguiendo la distribución de la Tabla 3. Estas muestras se prepararon para cada uno de los pimentones seleccionados como representantes de todo el conjunto inicial, lo que hace un total de 54 muestras de calibrado. En el conjunto de calibrado,

cada colorante se encuentra en cinco niveles de concentración distintos que equivalen a un porcentaje de colorante en pimentón entre 0 y 3,2% y además se tienen 6 matrices distintas de pimentón. La Figura 16 muestra los espectros correspondientes a estas nueve muestras, en el caso del pimentón número 1.

**Tabla 3.** Concentración de los colorantes en el extracto.

	E-102 (mg L <sup>-1</sup> )	E-110 (mg L <sup>-1</sup> )
1	0,8	0,4
2	0,6	0,2
3	0,4	0
4	0,2	0,2
5	0	0,4
6	0,2	0,6
7	0,4	0,8
8	0,6	0,6
9	0,4	0,4



**Figura 16.** Espectros correspondientes a las 9 muestras del pimentón 1 adulteradas con colorantes artificiales.

Por otro lado se escogieron otros 6 pimentones, distintos de los que constituyen el conjunto de calibrado, y se adulteraron con cada uno de los colorantes a dos niveles de concentración. Estas muestras constituyen el conjunto test y servirán para evaluar la capacidad de predicción de los modelos de regresión en pimentones ajenos al conjunto de calibrado. Los pimentones que formaron parte del conjunto test fueron 10, 14, 20, 24, 31, 35 y las concentraciones de dicho conjunto se muestran en la Tabla 4.

**Tabla 4.** Concentración de los colorantes en el extracto de las muestras test.

	E-102 (mg L <sup>-1</sup> )	E-110 (mg L <sup>-1</sup> )
10.1	0,3	0,5
10.2	0,5	0,3
14.1	0,5	0,5
14.2	0,3	0,3
20.1	0,2	0,6
20.2	0,6	0,2
24.1	0,7	0,7
24.2	0,1	0,1
31.1	0,1	0,7
31.2	0,7	0,1
35.1	0,2	0,2
35.2	0,6	0,6

En primer lugar, se realizó el análisis de componentes principales con las 54 muestras de calibrado. Con las dos primeras componentes principales se explica un 82,86% de varianza. No se puede apreciar un agrupamiento en cuanto a las concentraciones añadidas de cada adulterante, las muestras se codifican con el número de pimentón y el nivel que se especifica en la Tabla 3.

### **5.2.2 Modelos de calibrado a partir de un único pimentón**

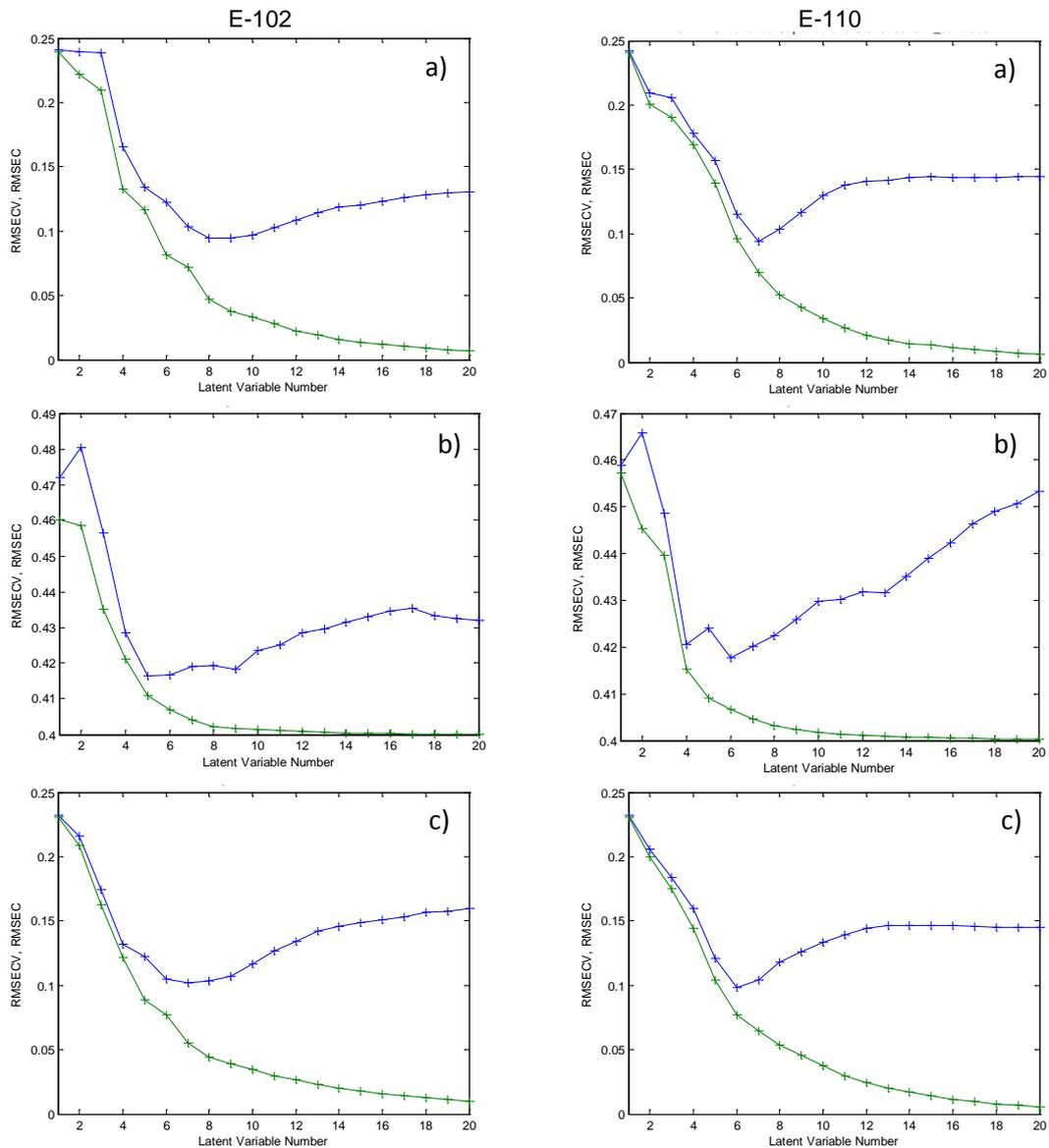
La primera etapa en la construcción de los modelos de calibrado PLS para cada adulterante ha sido el estudio del pretratamiento de los datos más adecuado. Para ello, se probó entre distintos pretratamientos y se eligió aquel que conducía a modelos con una menor diferencia entre el error cuadrático medio de calibración (RMSEC) y el error cuadrático medio de la validación cruzada (RMSECV), ya que cuanto menor sea la diferencia entre ambos errores, mayor capacidad de predicción tiene el modelo de calibrado.

En la Figura 17 se muestran RMSEC y RMSECV para ambos adulterantes en función del número de variables latentes del modelo cuando éste se ha construido con los datos originales, con los datos estandarizados y con SNV (Standard Normal Variate). El método de validación cruzada, en todos los casos, fue el de persianas venecianas.

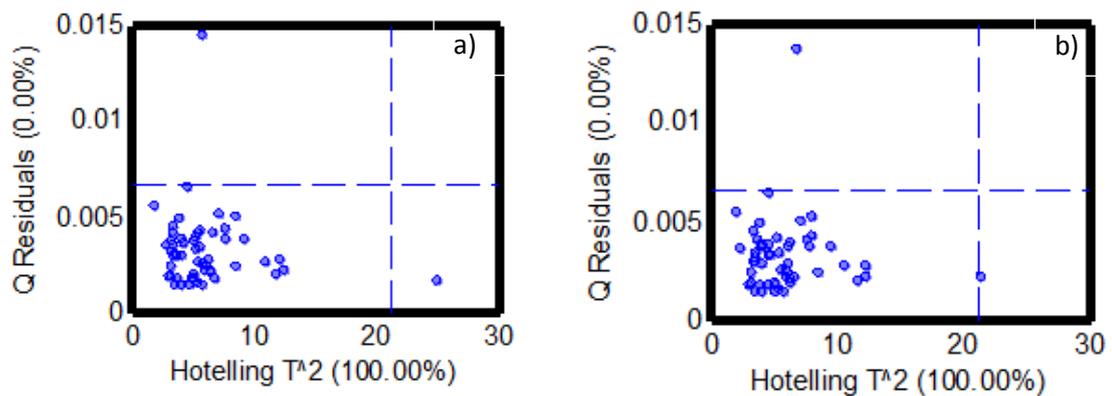
Para los dos adulterantes, se ha elegido un modelo PLS con los datos originales y con 6 variables latentes que proporciona errores más bajos y menor diferencia entre los errores cuadráticos medios.

Para analizar la posible presencia de muestras anómalas se realiza un estudio de los índices Q y  $T^2$  de Hotelling de los objetos, a un nivel de confianza del 99%. Para una muestra, un valor alto de estos estadísticos, por encima de los valores umbral para ambos índices, indica que ésta se encuentra fuera del hiperplano formado por el conjunto de muestras (Q alto) y alejada del centroide del mismo ( $T^2$  de Hotelling alto). Por tanto, esa muestra debe ser eliminada porque hay indicios para asegurar que es anómala respecto del resto.

En este caso para ninguno de los dos adulterantes ha sido necesario eliminar muestras, ya que no se ha obtenido ningún punto anómalo, tal y como se puede observar en la Figura 18.



**Figura 17.** RMSEC (en verde) y RMSECV (en azul) con (a) datos originales, (b) datos estandarizados y (c) con SNV para ambos adulterantes.



**Figura 18.** Índices Q y  $T^2$  de Hotelling (valores umbral a un nivel de confianza del 99%) obtenidos en el análisis del modelo de calibrado para (a) tartrazina y (b) amarillo anaranjado.

### 5.2.3 Modelos de calibrado conjuntos

Los modelos de calibrado obtenidos con un único tipo de pimentón en el conjunto de calibrado fueron utilizados para predecir la concentración de los adulterantes en el conjunto test y se evaluó el error obtenido en esas muestras a través del error cuadrático medio en predicción (RMSEP). Se observó que el modelo únicamente era capaz de predecir correctamente las muestras correspondientes al pimentón que se había utilizado en el conjunto de calibrado.

Por este motivo, y con el objetivo de incrementar la capacidad predictiva del modelo, al conjunto de calibración inicial, se le fueron añadiendo sucesivamente conjuntos de 9 pimentones de calibrado correspondientes a nuevos pimentones. Con cada modelo construido al adicionar un nuevo pimentón al calibrado se evaluó el RMSEP en el conjunto test. De esta forma se obtuvo un valor de error en predicción cada vez más bajo al tener más muestras en el conjunto de calibrado y, por lo tanto, más información para modelar el efecto matriz de los distintos pimentones.

En la Tabla 5 se muestran el número de variables latentes escogidas para cada modelo, y los valores de los errores en calibrado, por validación cruzada y en predicción para el adulterante E-102. En la Tabla 6 se muestran los mismos valores pero en este caso para el adulterante E-110. Aparecen remarcados los valores del modelo seleccionado en cada caso, que es utilizado posteriormente para la predicción del conjunto test.

En el caso del E-110 el error en predicción no disminuye por el hecho de añadir otros pimentones al calibrado. Sin embargo, en el caso del E-102, el RMSEP disminuye al aumentar el número de pimentones en el calibrado; es decir, cuantas más muestras tiene el calibrado mejor modela el efecto matriz de los distintos pimentones y predice las concentraciones de las muestras test.

**Tabla 5.** Variables latentes, RMSEC, RMSECV y RMSEP para el E-102.

	6 pimentones		7 pimentones		8 pimentones		9 pimentones		10 pimentones		11 pimentones	
<b>VL</b>	RMSEC	RMSECV	RMSEC	RMSECV	RMSEC	RMSECV	RMSEC	RMSECV	RMSEC	RMSECV	RMSEC	RMSECV
<b>1</b>	0,231	0,232	0,231	0,231	0,231	0,231	0,231	0,260	0,231	0,260	0,231	0,259
<b>2</b>	0,209	0,216	0,208	0,215	0,216	0,215	0,212	0,245	0,210	0,243	0,208	0,243
<b>3</b>	0,162	0,175	0,172	0,181	0,171	0,177	0,166	0,196	0,164	0,193	0,166	0,195
<b>4</b>	0,122	0,132	0,142	0,153	0,145	0,155	0,143	0,168	0,141	0,166	0,151	0,173
<b>5</b>	0,089	0,123	0,090	0,128	0,103	0,126	0,105	0,135	0,108	0,135	0,117	0,142
<b>6</b>	0,077	0,105	0,082	0,106	0,094	0,116	0,100	0,122	0,102	0,125	0,109	0,132
<b>7</b>	0,055	0,102	0,070	0,112	0,085	0,119	0,092	0,124	0,091	0,124	0,089	0,121
<b>8</b>	0,045	0,103	0,058	0,104	0,070	0,115	0,074	0,124	0,077	0,122	0,074	0,122
<b>9</b>	0,039	0,107	0,050	0,103	0,062	0,114	0,065	0,122	0,066	0,120	0,068	0,124
<b>10</b>	0,034	0,117	0,047	0,107	0,054	0,120	0,060	0,125	0,061	0,122	0,066	0,124
<b>RMSEP</b>	<b>0,345</b>		<b>0,374</b>		<b>0,308</b>		<b>0,310</b>		<b>0,275</b>		<b>0,244</b>	

**Tabla 6.** Variables latentes, RMSEC, RMSECV y RMSEP para el E-110.

	6 pimentones		7 pimentones		8 pimentones		9 pimentones		10 pimentones		11 pimentones	
VL	RMSEC	RMSECV	RMSEC	RMSECV	RMSEC	RMSECV	RMSEC	RMSECV	RMSEC	RMSECV	RMSEC	RMSECV
1	0,231	0,232	0,231	0,231	0,231	0,231	0,231	0,260	0,231	0,260	0,231	0,260
2	0,200	0,206	0,196	0,200	0,209	0,212	0,205	0,234	0,203	0,232	0,204	0,232
3	0,175	0,184	0,172	0,179	0,181	0,188	0,178	0,211	0,178	0,210	0,175	0,210
4	0,144	0,159	0,141	0,154	0,142	0,152	0,148	0,189	0,148	0,191	0,141	0,181
5	0,104	0,121	0,106	0,119	0,112	0,125	0,134	0,166	0,132	0,166	0,129	0,158
6	0,077	0,098	0,085	0,104	0,087	0,105	0,107	0,136	0,109	0,138	0,112	0,136
7	0,064	0,104	0,074	0,101	0,072	0,101	0,096	0,133	0,096	0,132	0,096	0,128
8	0,054	0,118	0,066	0,106	0,069	0,103	0,088	0,132	0,090	0,131	0,092	0,132
9	0,046	0,126	0,054	0,114	0,053	0,109	0,069	0,151	0,073	0,146	0,077	0,138
10	0,037	0,133	0,048	0,116	0,047	0,113	0,059	0,153	0,064	0,147	0,066	0,139
RMSEP	0,174		0,183		0,174		0,174		0,174		0,174	

Por último, para determinar la concentración de los colorantes en pimentón se utilizó el modelo de regresión PLS que proporciona menores valores de RMSEP. En el caso del E-102, como conjunto de calibrado se tiene un total de 99 muestras, correspondientes a extractos acuosos de 11 pimentones distintos, donde el E-102 varió entre 5 niveles de concentración. El modelo de regresión PLS con los datos originales y con 6 variables latentes explica un 95% de varianza en la respuesta (Tabla 7). Este modelo se utilizó para predecir 12 muestras test cuya concentración se muestra en la Tabla 4.

Para la determinación del E-110, el conjunto de calibrado estaba formado por 72 extractos de 8 pimentones distintos en los que el E-110 variaba en 5 niveles de concentración. El modelo de regresión PLS con los datos originales y con 6 variables latentes explica 97% de varianza en la respuesta (Tabla 7). Este modelo se utilizó para predecir 39 muestras test compuestas por el extracto acuoso de otros 9 pimentones.

**Tabla 7.** Modelos de regresión PLS para la determinación de los adulterantes.

	Variables latentes	Varianza acumulada X (%)	Varianza acumulada Y (%)	R <sup>2</sup> cal	E <sub>rel,cal</sub> (%)
<b>E-102</b>	6	99,99	94,57	0,75	21,74
<b>E-110</b>	6	99,99	96,86	0,87	24,72

En la Figura 19 se muestran los errores relativos alcanzados en la predicción de las muestras de calibrado frente a la concentración de los dos colorantes. Se puede observar como a medida que disminuye la concentración de las muestras se incrementa la magnitud de los errores relativos hasta alcanzar valores que no son asumibles. Esto indica que este método de análisis no sería válido para realizar determinaciones hasta esos niveles de concentración.

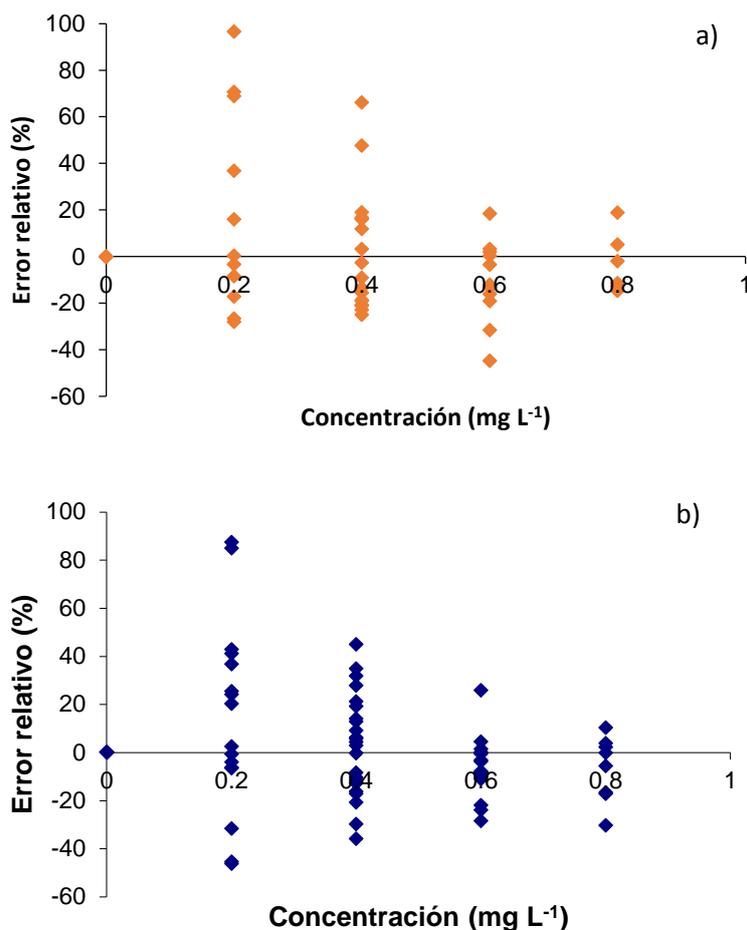


Figura 19. Errores relativos predichos frente a la concentración de (a) E-102 y (b) E-110.

## 6. Conclusiones

El análisis de componentes principales permite diferenciar los pimentones que no son “de la Vera” del resto, pero no permite clasificarlos según el tipo de pimentón. Los mismos resultados se obtienen cuando se aplica este análisis a las variables CIELab medidas en esas muestras de pimentón.

El análisis cluster llevado a cabo, tanto con los datos espectrales como con las variables CIELab, conduce a resultados similares en cuanto al número de grupos de pimentones obtenidos cuando se utiliza el método de Ward de agrupamientos. En este caso los grupos obtenidos no se pueden relacionar ni con el tipo ni con el origen del pimentón.

En cuanto a la determinación de la concentración de adulterante (E-102 y E-110) en el extracto acuoso de las muestras de pimentón, la regresión de mínimos cuadrados parciales permite confirmar la presencia de los adulterantes, pero los errores obtenidos son elevados. Esto probablemente se deba a que los niveles de adulteración más bajos considerados en este trabajo se encuentran en el límite de detección del procedimiento de análisis utilizado, sería necesario concentrar más los extractos para reducir dichos errores.

## 7. Bibliografía

1. A. Palacios. Tesis doctoral: *Caracterización analítica y diferenciación geográfica de pimentón mediante técnicas de reconocimiento de patrones*. Universidad de Sevilla, 2015.
2. F. Márkus, H.G. Dado, J. Kapitány and 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*, 47 (1): 100-107, 1999.
3. O.C. Zúñiga, F.G. Jiménez and R.M. Gordillo. "Comparative study of carotenoid composition in three Mexican varieties of *Capsicum annum L.*" *Food Chemistry*, 90: 109-114, 2005.
4. I. Mínguez, *Carotenoides en el pimentón. Factores responsables de su degradación*. CSIC 1996.
5. Consejo Regulador de la D.O. "Pimentón de la Vera", Denominación de Origen Protegida "Pimentón de la Vera", Pliego de condiciones (Cáceres). <https://pimentonvera-origen.com/> (último acceso 05/02/2019).
6. Orden de 16 de diciembre de 1999, de la Consejería de Agricultura, Agua y Medio Ambiente, por la que se concede una protección transitoria a la denominación de origen Pimentón de Murcia y se crea su consejo regulador. B.O. de la Región de Murcia, nº 295, p 13231.
7. AENOR. *Pimentón (Capsicum annum L.). Especificaciones*. UNE EN-ISO 75-40. Madrid: AENOR. 2010.
8. K.M.S Gomes, M.V.G.A.D Oliveira, F.R.D.S. Carvalho, C.C Menezes and A.P. Peron. "Citotoxicity of food dyes sunset yellow (E-110), bordeaux red (E-123), and tatzarine yellow (E-102) on *Allium cepa L.* root meristematic cells". *Food Science and Technology*, 33 (1): 218-223, 2013.
9. I. Emadfa, E. Mustak and D. Fritzsche. *Tabla de aditivos los números E*. Madrid. Hispano Europea. 2011.
10. Reglamento (CE) nº 1333/2008 del Parlamento Europeo y del Consejo sobre aditivos alimentarios. DOUE 354 del 31/12/2008, 16-33.
11. Reglamento (UE) nº 1129/2011 de la Comisión por el que se modifica el anexo II del Reglamento (UE) nº 1333/2008 del Parlamento Europeo y del Consejo para establecer una lista de aditivos alimentarios de la Unión. DOUE 295 del 12/11/2011, 1-177.
12. ASTA, *Official analytical methods of the American spice trade association* (2<sup>nd</sup> ed.), Englewood Cliffs: ASTA, 1986.
13. A. Pérez-Gálvez, M.I. Mínguez-Mosquera, J. Garrido-Fernández, M. Lozano-Ruiz and V. Montero-de-Espinosa. "Correlation between ASTA units-carotenoid concentration in paprika. Prediction of the color stability during storage". *Grasas y Aceites*, 55: 213-218, 2004.

14. CIE, 15 technical report: Colorimetry (3<sup>rd</sup> ed.), 2004.
15. A. Kaur and B.V. Kranthi. "Comparison between YCbCr color space and CIELab color space for skin color segmentation". *International Journal of Applied Information Systems*, 3 (4): 30-33, 2012
16. M. Casale, P. Oliveri, C. Armanino, S. Lanteri and M. Forina. "NIR and UV-vis spectroscopy, artificial nose and tongue: Comparison of four fingerprinting techniques for the characterisation of Italian red wines". *Analytica Chimica Acta*, 668 (2): 143-148, 2010.
17. M. Urbano, M.D. Luque de Castro, P.M. Pérez, J. García-Olmo and M.A. Gómez-Nieto. "Ultraviolet-visible spectroscopy and pattern recognition methods for differentiation and classification of wines". *Food Chemistry*, 97 (1): 166-175, 2006.
18. U.T.C.P Souto, M.J.C. Pontes, E.C. Silva, R.K.H. Galvão, M.C.U. Araújo and F.A.C. Sanches et al. "UV-Vis spectrometric classification of coffees by SPA-LDA". *Food Chemistry*, 119: 368-371, 2010.
19. P.H.G.D Diniz, M.F. Barbosa, K.D.T.M. Milanez, M.F. Pistonesi and M.C.U. Araujo. "Using UV-Vis spectroscopy for simultaneous geographical and varietal classification of tea infusions simulating a home-made tea cup". *Food Chemistry*, 192: 374-379, 2016.
20. P. Lin, Y. Chen and Y. He. "Identification of geographical origin of olive oil using visible and near-infrared spectroscopy technique combined with chemometrics". *Food and Bioprocess Technology*, 5 (1): 235-242, 2012.
21. M. Brunner, R. Katona, Z. Stefánka and T. Prohaska. "Determination of the geographical origin of processed spice using multielement and isotopic pattern on the example of Szegedi paprika". *European Food Research and Technology*, 231: 623-634, 2010.
22. A. Palacios-Morillo, J.M. Jurado, A. Alcázar and F. Pablos. "Geographical characterization of Spanish PDO paprika by multivariate analysis of multielemental content". *Talanta*, 128: 15-22, 2014.
23. A. Palacios-Morillo, J.M. Jurado, A. Alcázar and F. Pablos. "Differentiation of Spanish paprika from Protected Designation of Origin based on color measurements and pattern recognition". *Food Control*, 62: 243-249, 2016.
24. A. Sánchez. *Uso de técnicas quimiométricas para clasificar pimentones utilizando espectros de absorción molecular y variables CIELab*. Trabajo fin de grado. Burgos: Universidad de Burgos, 2016.
25. I. Lee and J. Yang. *Common Clustering Algorithms*. En *Comprehensive Chemometrics Chemical and Biochemical Data Analysis* (pp. 577-618). Oxford: Elsevier, 2009
26. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. Jong and P.J. Smeyers-Verbeke. *Handbook of Chemometrics and Qualimetrics. Parte A, en Data Handling in Science and Technology*. Elsevier, Amsterdam, 1997.

27. H. Martens and T. Naes. Multivariate calibration, John Wiley & Sons, Nueva York, 1989.
28. Statgraphics Centurion XVII. StatPoint Technologies, Inc., Herndon, VA, 2014.
29. MATLAB versión 8.4.0.150421 (R2014b), The Mathworks, Inc., Natick, MA, USA, 2014.