



Aplicación de técnicas de imagen  
para la evaluación de la madurez  
fenólica de semillas de uva

**Francisco José Rodríguez Pulido**  
Sevilla, octubre de 2013



**Dña. Ana María Cameán Fernández**, Directora del Departamento de Nutrición y Bromatología, Toxicología y Medicina Legal de la Universidad de Sevilla y **D. José Manuel Vega Pérez**, Coordinador del Programa de Doctorado en Farmacia,

**CERTIFICAN:**

Que el presente trabajo ha sido realizado dentro del Programa de Doctorado en Química por el Licenciado en Química D. Francisco José Rodríguez Pulido, siendo sus directores D. Francisco José Heredia Mira y Dña. María Lourdes González-Miret Martín, habiéndose cumplido todos los requisitos legales vigentes,

Y para que conste, se expide y firma el presente certificado en Sevilla, a 8 de octubre de 2013.



Fdo. Ana María Cameán Fernández



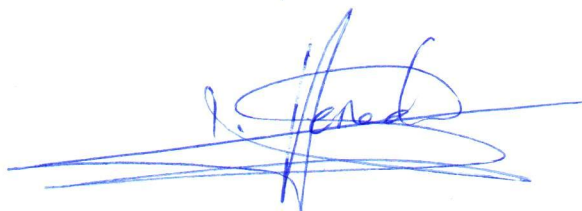
Fdo. José Manuel Vega Pérez





La tesis titulada ***“Aplicación de técnicas de imagen para la evaluación de la madurez fenólica de semillas de uva”***, presentada por el Licenciado en Química **D. Francisco José Rodríguez Pulido** para optar al grado de Doctor, ha sido realizada en el Área de Nutrición y Bromatología del Departamento de Nutrición y Bromatología, Toxicología y Medicina Legal (Facultad de Farmacia, Universidad de Sevilla), bajo la dirección de los Doctores **D. Francisco José Heredia Mira**, Catedrático de la Universidad de Sevilla y **Dña. María Lourdes González-Miret Martín**, Profesora Titular de la Universidad de Sevilla. Considerando que se han alcanzado los objetivos inicialmente previstos, y que el trabajo reúne los requisitos necesarios, autorizan su presentación.

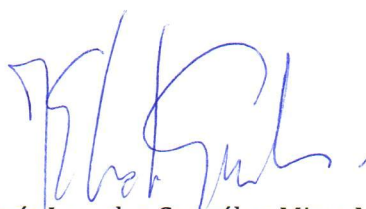
En Sevilla, a 9 de octubre de 2013

A blue ink signature in cursive script, appearing to read 'F. Heredia Mira'.

Fdo. Francisco José Heredia Mira

A blue ink signature in cursive script, appearing to read 'F. Rodríguez Pulido'.

Fdo. Francisco José Rodríguez Pulido

A blue ink signature in cursive script, appearing to read 'M. Lourdes González-Miret Martín'.

Fdo. María Lourdes González-Miret Martín

Directores de la Tesis



Este trabajo ha sido realizado con el apoyo institucional del Ministerio de Economía y Competitividad del Gobierno de España con el proyecto “Aplicación de la colorimetría triestímulo al estudio del papel de los flavanoles en la copigmentación con antocianos” (AGL2008-05569-C02-02), del que el doctor Francisco José Heredia Mira es investigador principal, y la Beca Predoctoral (BES-2009-025429) asociada a dicho proyecto. Parte de este trabajo ha sido desarrollado también con el apoyo del Ministerio de Economía y Competitividad con la ayuda para la realización de una estancia breve en Dublín de tres meses (EEBB-2011-43931), gracias a la cual ha sido posible optar a la Mención Internacional del Título de Doctor.





*A mis hermanos,  
Joaquín, Lola y Alicia*

*«La mayoría de las ideas fundamentales de la ciencia son esencialmente sencillas y,  
por regla general pueden ser expresadas en un lenguaje comprensible para todos.»*

Albert Einstein



---

## ÍNDICE

<b>PRODUCCIÓN CIENTÍFICA</b>	<b>5</b>
<b>RESUMEN</b>	<b>7</b>
<b>INTRODUCCIÓN</b>	<b>13</b>
<b>1. LAS SEMILLAS DE UVA</b>	<b>15</b>
1.1. Composición	16
1.1.1. Flavonoides	18
1.1.2. No flavonoides	21
1.2. Maduración	21
1.2.1. Aspectos generales y desarrollo de la uva	22
1.2.2. Problemática de los climas cálidos	23
1.2.3. Desarrollo fisiológico de la semilla de uva	24
<b>2. EL COLOR</b>	<b>29</b>
2.1. Espacio triestímulo	32
2.2. Ecuaciones colorimétricas básicas	34
2.3. Sistema colorimétrico CIE	35
2.4. Sistemas CIE 1931(x,y) y CIE 1964-(x10,y10) (CIEXYZ)	37
2.5. Espacio CIELAB	39
<b>3. VISIÓN ARTIFICIAL</b>	<b>43</b>
3.1. Fundamentos de imágenes digitales	46
3.1.1. Resolución	47
3.1.2. Profundidad de bits	47
3.2. Adquisición de imágenes	48
3.3. Segmentación	49
3.3.1. Técnicas de segmentación	51
3.3.2. Transformaciones morfológicas	52
3.4. Análisis morfológico de imágenes digitales	53
3.5. Análisis colorimétrico de imágenes digitales	55
3.5.1. Obtención de las coordenadas CIELAB a partir del espacio RGB	57
3.6. Análisis de imágenes hiperespectrales	61
3.6.1. Fundamentos de espectroscopía	64
3.6.2. Interacción de la luz con la muestra	66
3.6.3. Terminología	67
3.6.4. Tratamientos espectrales	69
<b>OBJETIVOS GENERALES</b>	<b>73</b>

---

<b>MATERIALES Y MÉTODOS</b>	<b>79</b>
<b>1. MUESTRAS</b>	<b>81</b>
1.1. Proceso de muestreo	84
<b>2. ANÁLISIS DE IMAGEN</b>	<b>85</b>
2.1. Adquisición de imágenes digitales	85
2.1.1. Cabina de iluminación	86
2.1.2. Cámara digital	86
2.2. Adquisición de imágenes hiperespectrales	87
2.2.1. Cámara	88
2.2.2. Espectrógrafo	88
2.2.3. Sistema de barrido	88
2.2.4. Iluminación	89
2.2.5. Bandeja de muestra	89
2.2.6. Blanco de referencia	89
2.2.7. Software para la adquisición de imágenes	89
<b>3. ANÁLISIS QUÍMICO</b>	<b>91</b>
3.1. Preparación de las muestras	91
3.1.1. Material y reactivos	91
3.1.2. Procedimiento analítico	92
3.2. Análisis cromatográfico	93
3.2.1. Material y reactivos	93
3.2.2. Procedimiento analítico	94
3.3. Análisis espectrofotométrico	94
3.3.1. Material y reactivos	94
3.3.2. Procedimiento analítico	95
<b>4. TRATAMIENTO DE RESULTADOS</b>	<b>97</b>
4.1. Procesado de imágenes	97
4.2. Análisis estadísticos y quimiométricos	98
4.2.1. Análisis de la varianza (ANOVA)	99
4.2.2. Análisis General Discriminante (GDA)	99
4.2.3. Análisis de Componentes Principales (PCA)	100
4.2.4. Regresión Múltiple	102
<b>RESULTADOS Y DISCUSIÓN</b>	<b>105</b>
<b>CAPÍTULO 1. Metodología para la adquisición de imágenes digitales de las uvas y sus semillas. Estudio de la evolución colorimétrica a lo largo de la maduración</b>	<b>107</b>
1.1. Antecedentes	107
1.2. Objetivos	109
1.3. Parte experimental	110
1.4. Conclusiones	114
• <i>Publicación:</i> "Ripeness estimation of grape berries and seeds by image analysis". Computers and Electronics in Agriculture, 82 (2012), 128-133.	117

---

<b>CAPÍTULO 2. Evaluación del color mediante técnicas de análisis de imagen y establecimiento de la metodología para la cuantificación de nubes de puntos</b>	<b>123</b>
2.1. Antecedentes	123
2.2. Objetivos	124
2.3. Parte experimental	125
2.3.1. Construcción de los elipsoides y evaluación de la capacidad predictiva	125
2.3.2. Propuesta de un índice de maduración de semillas basado en sus características colorimétricas	128
2.4. Conclusiones	132
• <i>Publicación</i> : “Analysis of food appearance properties by computer vision applying ellipsoids to colour data”.	
Computers and Electronics in Agriculture, (2013), 99, 108-115.	135
<b>CAPÍTULO 3. Relación entre el color y la composición fenólica de semillas de uva</b>	<b>143</b>
3.1. Antecedentes	143
3.2. Objetivos	144
3.3. Parte experimental	145
3.3.1. Estudio de la relación apariencia-composición mediante estadística univariante	145
3.3.2. Caracterización del perfil fenólico y uso de técnicas multivariantes para la estimación de la etapa de madurez	146
3.4. Conclusiones	148
• <i>Publicación</i> : “Application of Image Analysis to the Color-Phenolic Composition Relationships of Grape Seeds”. En: José Luis Caivano y María del Pilar Buera (Eds.) <i>Color in Food. Technological and Psychophysical Aspects</i> , 113 -120. (2012), CRC Press.149	
• <i>Publicación</i> : “Preliminary study to determine the phenolic maturity stage of grape seeds by computer vision”. <i>Analytica Chimica Acta</i> , (2012), 732, 78-82.	159
<b>CAPÍTULO 4. Aplicación de las técnicas hiperespectrales a las semillas de uva. Establecimiento de la metodología y discriminación varietal</b>	<b>165</b>
4.1. Antecedentes	165
4.2. Objetivos	166
4.3. Parte experimental	166
4.4. Conclusiones	171
• <i>Publicación</i> : “Grape seed characterization by NIR hyperspectral imaging”.	
Postharvest Biology and Technology, (2013), 76, 74-82.	173
<b>CAPÍTULO 5. Estudio de la relación entre la composición química y el perfil hiperespectral en las semillas de uva</b>	<b>183</b>
5.1. Antecedentes	183
5.2. Objetivos	184
5.3. Parte experimental	184
5.3.1. Análisis de referencia	185
5.3.2. Adquisición y procesado de imágenes hiperespectrales	186
5.3.3. Análisis exploratorio y cuantitativo	188
5.4. Conclusiones	189
• <i>Manuscrito</i> : A novel method for evaluating flavanols in grape seeds by near infrared hyperspectral imaging. <i>Talanta (en revisión)</i>	191

---

<b>CONCLUSIONES GENERALES</b>	<b>215</b>
<b>REFERENCIAS</b>	<b>223</b>
<b>ANEXO</b>	<b>237</b>

---

## PRODUCCIÓN CIENTÍFICA

La estructura de esta Memoria sigue la modalidad "Conjunto de Trabajos" recogida en el Artículo 9 de la Normativa Reguladora de Tesis Doctoral de la Universidad de Sevilla (BOUS nº3 de 23 de mayo de 2012) a partir de la siguiente producción científica:

### Artículos:

- Francisco J. Rodríguez-Pulido, Luis Gómez-Robledo, Manuel Melgosa, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia, (2012). *Ripeness estimation of grape berries and seeds by image analysis*, Computers and Electronics in Agriculture, 82, 128-133, ISSN 0168-1699.
- Francisco J. Rodríguez-Pulido, Raúl Ferrer-Gallego, M. Lourdes González-Miret, Julián Carlos Rivas-Gonzalo, María Teresa Escribano-Bailón, Francisco J. Heredia, (2012). *Preliminary study to determine the phenolic maturity stage of grape seeds by computer vision*, Analytica Chimica Acta, 732, 78-82, ISSN 0003-2670.
- Francisco J. Rodríguez-Pulido, Douglas F. Barbin, Da-Wen Sun, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia, (2013). *Grape seed characterization by NIR hyperspectral imaging*, Postharvest Biology and Technology, 76, 74-82, ISSN 0925-5214.
- Francisco J. Rodríguez-Pulido, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia, (2013). *Analysis of food appearance properties by computer vision applying ellipsoids to colour data*, Computers and Electronics in Agriculture, 99, 108-115, ISSN 0168-1699.
- Francisco J. Rodríguez-Pulido, José Miguel Hernández-Hierro, Julio Nogales-Bueno, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia. *A novel method for evaluating flavanols in grape seeds by near infrared hyperspectral imaging*. Talanta, ISSN: 0039-9140. (En revisión)

### Capítulo de Libro:

- Francisco José Rodríguez-Pulido, Francisco J. Heredia, Juan Manuel Zaldívar-Cruz, M. Lourdes González-Miret, (2012). *Application of Image Analysis to the Color-Phenolic Composition Relationships of Grape Seeds*. En: José Luis Caivano y María del Pilar Buera (Eds.) *Color in Food. Technological and Psychophysical Aspects*, 113 -120. CRC Press. ISBN 978-1-4398-7693-0

### Patentes y Registros de la Propiedad Intelectual:

- Patente de dispositivo para la evaluación de la madurez de semillas de uva mediante digitalización de imágenes (*en trámite*)
- *Software para el análisis morfológico y colorimétrico de las semillas de uva (DigiSeed®)*. Registro territorial de la Propiedad Intelectual de Andalucía. (Expediente SE-471-12)
- *Software para el análisis morfológico y colorimétrico de uvas (DigiGrape®)*. Registro territorial de la Propiedad Intelectual de Andalucía. (Expediente SE-472-12)





# *RESUMEN*





---

## RESUMEN

La calidad del vino depende ampliamente del estado de todos los componentes de la uva, y en este aspecto, sus semillas afectan directamente a la estructura, astringencia, estabilidad, e indirectamente al color del vino. Los compuestos fenólicos que contienen son responsables de estas propiedades, y cambian de manera cualitativa y cuantitativa durante toda la maduración del fruto. Además de estar implicados en la calidad del vino, poseen un alto valor farmacológico debido a su capacidad antioxidante.

El desarrollo de la uva sucede en tres etapas principales: crecimiento herbáceo, maduración y sobremaduración. Es durante la maduración cuando se produce en el hollejo la pérdida de clorofila y la aparición de los colorantes definitivos. Mientras, tanto en los hollejos como en las semillas, los fenoles polimerizan disminuyendo su astringencia. La polimerización de fenoles y la disminución de la acidez del mosto disminuyen la agresividad de los taninos en las bayas, y afectan a las cualidades gustativas de las distintas partes de la uva. Al mismo tiempo, todos estos procesos se ven afectados por factores agronómicos como el clima, el suelo, la disponibilidad de agua, la fertilización, la variedad, etc. Por ejemplo, en climas cálidos, las primaveras secas y los veranos calurosos hacen que el periodo comprendido entre el envero y la madurez industrial de la uva sea más corto. Así, mientras que la pulpa ha alcanzado niveles óptimos de azúcar, las semillas aún no se han desarrollado plenamente, lo que produce desequilibrios en el vino producidos por la falta de madurez fenólica de las semillas de uva. Por estos motivos es tan importante el control del estado de maduración de las semillas de uva. En el presente trabajo se han evaluado las semillas de uva de diferentes variedades aplicando diferentes técnicas de análisis de imagen. En todos los casos se ha intentado el desarrollo de metodologías que implicaran la mínima preparación de muestra y la máxima rapidez del método, con el objeto de poder sustituir en la medida de lo posible el análisis químico, que implica largos y tediosos procesos que impiden un seguimiento rápido de la maduración.

---

Se ha aplicado la colorimetría triestímulo mediante técnicas de imagen, que permitió la discriminación varietal de semillas de uva a partir de sus características colorimétricas y morfológicas, y permitió el estudio de su evolución durante la maduración. También se ha desarrollado un método para la cuantificación de la heterogeneidad del color en alimentos. La metodología propuesta para la medición y evaluación del color ha sido satisfactoria para predecir el estado de madurez de las semillas de uva, y ha sido útil para establecer etapas de maduración a partir de sus características de apariencia.

Más allá del color, también se han evaluado las semillas de uva en el intervalo infrarrojo cercano del espectro mediante análisis de imágenes hiperespectrales. A partir de ellas, las semillas se han podido clasificar de manera más precisa que con el espectro visible, pudiendo incluso detectar diferencias en factores agronómicos a partir de esos espectros de reflectancia. La metodología puesta a punto en análisis de imágenes hiperespectrales resultó ser capaz de evaluar el contenido en flavanoles en semillas de uva mediante un sencillo análisis en cuestión de minutos y sin necesidad de ningún tratamiento de muestra.

---

## ABSTRACT

Quality of wine depends extensively on the condition of all the components of grapes. In this respect, their seeds directly affect to structure, astringency, stability, and indirectly to colour of wine. The phenolic compounds that they contain are responsible of these properties, and they change qualitatively and quantitatively along the berry ripening. Besides of their implications on quality of wine, they have a high pharmacological value due to their antioxidant properties.

The development of grapes happens in three main stages: growth, ripening, and over-ripening. The loss of chlorophylls and the development of final colorants occur during the ripening. Meanwhile, in skins as well as in seeds, phenols polymerise decreasing their astringency. The polymerisation of phenols and the loss of acidity in the grape-juice make the aggressiveness of tannins in berries decrease, and they affect to the gustatory properties of the different parts of the grape. At the same time, all these processes are affected by agronomical factors such as climate, soil, water availability, fertilisation, variety, etc. For instance, in warm climates, dry springs and hot summers make that the period between veraison and industrial maturity of grapes decrease. This way, while the flesh has reached optimal levels of sugar, the seeds have not developed yet thoroughly. This produce unbalances in wines due to the lack of ripeness of phenols in grape seeds. For these reasons it is so important the control of the state of ripening in grape seeds. In this work, grape seeds have been evaluated by several techniques of image analysis. In all cases, the methodology was developed trying to entail the minimum of sample preparation and the maximum of quickness of the method, and as far as possible, substitute the chemical analysis that involves long and tedious processes that hamper monitoring of ripening in a rapid way.

The tristimulus colorimetry has been applied by means of imaging techniques, that allowed the varietal discrimination of grape seeds from their colorimetric and morphological characteristics, and allowed the study of their evolution during the ripening. It was also developed a method for quantifying the heterogeneity of colour in food products. The proposed methodology for

---

measuring and evaluating colour has been satisfactory for predicting the ripening of grape seeds, and it has been useful for establishing stages of ripeness from their appearance features.

Beyond the colour, the grape seeds have been evaluated in near infrared spectrum by hyperspectral imaging. From these images, the seeds could be classified in a more accurate manner than by means of the visible spectrum, detecting even differences in agronomical factors from those reflectance spectra. The developed method of hyperspectral imaging was able to evaluate the flavanols content in grape seeds with a simple analysis in a few minutes and without any sample treatment.

# *INTRODUCCIÓN*







---

## 1. LAS SEMILLAS DE UVA

El vino, del latín, Vinum, es la bebida alcohólica que se obtiene de la fermentación de la uva (*Vitis Vinifera*) exprimida. Existen evidencias de su elaboración desde la antigüedad. Los egipcios se lo atribuyen a Osiris, y los griegos a Dioniso (Baco). Por su parte, los hebreos afirman que fue Noé el primero en cultivar la vid, y el vino ocupa un lugar privilegiado en sus ritos y costumbres. En España, los primeros datos sobre la elaboración del vino se sitúan en el año 2200 a.C.. Fueron los Tartessos los que por entonces no sólo cultivaron la vid y elaboraron vino, sino que comercializaron con él fuera de España (Delgado, 1995). Hoy día, España es el tercer productor a nivel mundial después de Francia e Italia y el cultivo de la vid representa casi el 60% del área de cultivo en Europa (OIV, 2013b).

El clima es uno de los factores más influyentes para el cultivo de la vid, ya que presenta unas exigencias climáticas bien definidas. Es una planta exigente de calor y sensible a las heladas en invierno y primavera, no sólo para su desarrollo vegetativo, sino también para la maduración de sus frutos, que precisan una iluminación y temperaturas adecuadamente altas. Por estas circunstancias, la vid se ha cultivado tradicionalmente en climas no extremos de tipo mediterráneo, y su área general de cultivo corresponde a las dos zonas terrestres comprendidas entre los paralelos 30° y 50° latitud Norte, y los 30° y 40° latitud Sur. El cambio climático

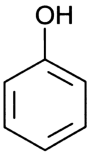
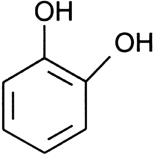
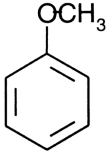
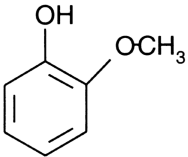
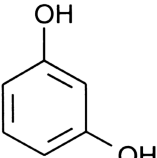
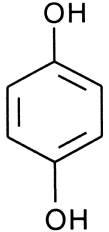
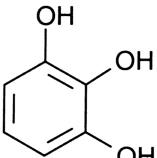
está provocando que estos límites se estén desplazando hacia latitudes mayores (Schultz & Jones, 2010).

### **1.1. Composición**

Los diferentes componentes de la uva han sido objeto de muchos estudios por su repercusión en la calidad final de un vino. Una familia de éstos son los compuestos fenólicos, responsables de la estructura, cuerpo, aroma y directamente implicados en el color del vino. No solamente influyen en la calidad del vino, también se ha demostrado que los compuestos fenólicos poseen un alto valor farmacológico por sus propiedades antioxidantes. De hecho, el consumo diario de las uvas y sus productos se asocia a un menor riesgo de padecer enfermedades degenerativas, como cardiovasculares y algunos tipos de cáncer (Xia et al., 2010). En las bayas de *Vitis Vinifera*, los fenoles se encuentran mayoritariamente en las partes sólidas (Cheynier & Rigaud, 1986), pasando al vino entre un 40 y un 60% durante el proceso prefermentativo (Zoecklein et al., 2001).

Los compuestos fenólicos incluyen todos aquellos que tienen al menos un anillo aromático con uno o más grupos hidroxilo, como el ácido cafeico. Por extensión, son compuestos polifenólicos los que poseen múltiples anillos fenólicos en su estructura, como son la catequina y el ácido elágico. Otra familia de compuestos importantes dentro de los fenoles son los taninos, término que describe algunas sustancias astringentes usadas para curtir pieles. Muchos compuestos fenólicos tienen esta propiedad, aunque tradicionalmente se denominan taninos a todo aquel extracto natural de plantas que contenga una mezcla compleja de compuestos fenólicos de alto peso molecular. Son polímeros de compuestos fenólicos más simples. La Tabla 1 muestra compuestos que a pesar de no estar presentes en la uva, son útiles para la comprensión de estructuras más complejas (Waterhouse, 2002).

Tabla 1. Compuestos fenólicos simples

Estructura	Nombre	Notas
	Fenol	Escaso en productos naturales, pero abundante en algunos aerosoles
	Pirocatecol, (1,2 dihidroxibenceno). Cuando está presente un grupo 1,2 dihidroxibenceno, se dice que es un grupo catecol	Fácilmente oxidado a o-quinona
	Anisol, (metoxibenceno)	
	Guayacol (2-metoxifenol). Este grupo es llamado guayacol cuando su estructura está presente.	La metilación dificulta la oxidación del grupo catecol
	Resorcinol (1,3 dihidroxibenceno)	No se oxida tan fácilmente como los dihidroxibencenos orto o para.
	1,4 dihidroquinona. (1,4 dihidroxibenceno)	También fácilmente oxidable a quinona, pero esta forma no es común en productos naturales
	Pirogallol 1,2,3-trihidróxido. Cuando un grupo 1,2,3-trihidróxido está presente, se dice que es un galol o un grupo galoilo	Fácilmente oxidable

Todos los compuestos fenólicos tienen mayor o menor susceptibilidad de ser oxidados y por ese motivo son reconocidos antioxidantes naturales. La reactividad de sus grupos hidroxilo hace que estén incluidos en las especies

reactivas de oxígeno (del inglés ROS, *reactive oxygen species*). Los compuestos que contienen 1,4-dihidroquinona, son especialmente fáciles de oxidar. Al hacerlo, el radical fenoxilo resultante puede ser estabilizado por el anión oxígeno adyacente. El radical formado es lo suficientemente estable y se mantendrá el tiempo suficiente para reaccionar con otro radical semiquinona. Esa reacción provocará una desproporción en quinona y fenol, y por consiguiente la extinción de dos radicales libres. El uso de sulfitos en la elaboración del vino mejora esta propiedad antioxidante mediante la reducción de la quinona, convirtiéndola de nuevo en fenol.

Todo los compuestos fenólicos se clasifican dos categorías, según contengan o no el esqueleto básico de los flavonoides:

### 1.1.1. Flavonoides

Los flavonoides son compuestos polifenólicos que comparten una estructura común (Figura 1). Esa estructura contiene varios anillos aromáticos con grupos hidroxilo.

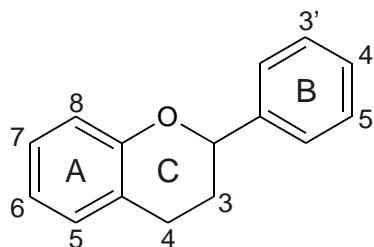


Figura 1. Esqueleto básico de un flavonoide

Las distintas clases de flavonoides dependen del estado de oxidación que tenga el anillo C. Por ejemplo, una cetona en el carbono 4 y una insaturación entre 2 y 3 define las flavonas; y el anillo C aromático, que implica una carga positiva sobre el oxígeno cíclico, define las antocianidinas. Esa carga positiva es especialmente estable al deslocalizarse en todo el sistema aromático. La mayoría de los fenoles en el vino tinto son flavonoides, y provienen de la extracción de los hollejos y las semillas durante el proceso de fermentación, ya que el medio

---

alcohólico favorece este proceso. Dentro de los flavonoides merece la pena destacar los flavanoles, flavonoles y antocianinas:

— *Flavanoles* (ej. catequina): Constituyen la clase más abundante, aunque la mayoría se encuentran formando oligómeros y polímeros en las semillas de uva. Son los flavonoides más abundantes en las uvas y el vino, y se han encontrado tanto en la semilla como en los hollejos. A veces se denominan flavan-3-oles para identificar la localización del grupo alcohol en el anillo C. Dado que las posiciones 2 y 3 del anillo están saturadas, existen esteroisómeros, y ambos se encuentran en las uvas de manera natural. La forma *trans* es (2R,3S) (+)-catequina y la forma *cis* es (2R,3R) (-)-epicatequina. Ambas formas tienen el grupo 3',4'-catecol en el anillo B.

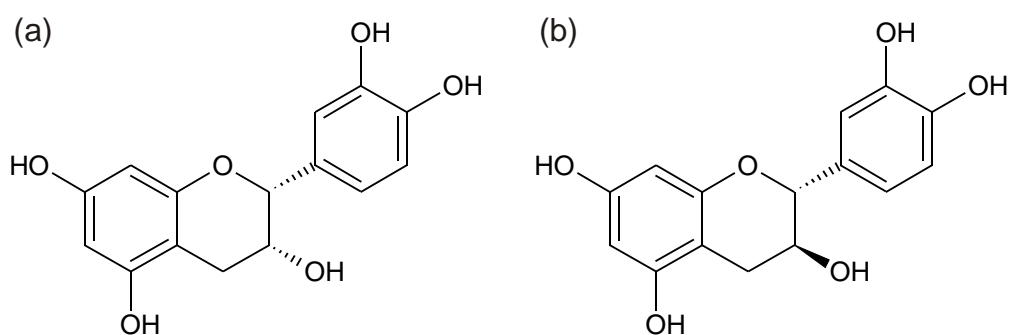


Figura 2. Formas *cis* y *trans* de los flavan-3-oles más comunes. (a) epicatequina y (b) catequina.

El anillo B puede tener grupos OH en los carbonos 3', 4' y 5', en cuyo caso se habla de galocatequinas. Algunas galocatequinas se encuentran en la piel de la uva, pero nunca en cantidades significativas.

- *Flavonoles* (ej. quercetina): Se encuentran mayoritariamente en la piel de la baya y funcionan a modo de protector solar, incrementándose su síntesis cuando la exposición a la radiación es mayor.
- *Antocianinas* (ej. malvidina-3-glucósido): Las antocianinas dan el color en el vino tinto y los colores rojo y azul oscuro en los hollejos de las uvas tintas. Este color se debe al sistema  $\pi$  de diez electrones conjugados entre los anillos A y C, incluso existe contribución del anillo B al estar en el mismo plano. El color se pierde si este sistema aromático se perturba, por ejemplo,

al tratar las antocianinas con bisulfito. Hay cinco antocianinas básicas en el vino: cianidina, peonidina, delphinidina, petunidina y malvidina (Figura 3).

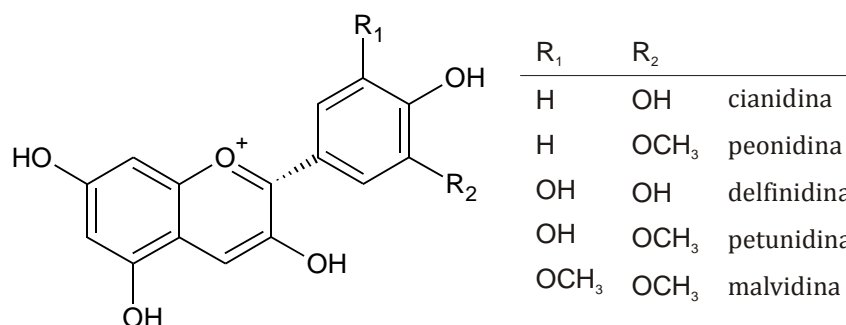


Figura 3. Estructura de antocianidinas comunes

Las antocianidinas interactúan con otros compuestos fenólicos en disolución para crear un fenómeno conocido como copigmentación (Boulton, 2001). Esta interacción es una unión no covalente y es el resultado del fenómeno químico de formación de complejos de transferencia de carga o interacciones  $\pi$ - $\pi$  (Foster, 1969). Esto sucede cuando dos sustancias en disolución con anillos aromáticos tienen densidades electrónicas muy diferentes. La carga positiva del catión flavilio hace a las antocianinas susceptibles de sufrir este fenómeno (Figura 4). La modificación del color que sufre el vino por la copigmentación ha sido estudiada en profundidad, no solo con medidas en las longitudes de onda clásicas donde se suelen cuantificar estos compuestos, sino considerando todo el espectro visible y evaluándose en un estudio colorimétrico integral (Gordillo et al., 2012).

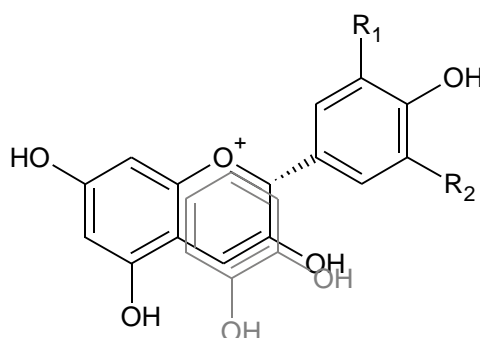


Figura 4. Reacción de copigmentación sobre el catión flavilio de las antocianinas

---

### 1.1.2. No flavonoides

El resto de compuestos fenólicos que no comparten esa estructura común se denominan genéricamente no flavonoides. Entre ellos podemos encontrar:

- *Ácidos hidroxicinámicos* (ej. ácido cafeico): Están presentes también en frutas y todos los tejidos vegetales, y se hallaron en forma de ésteres tartrato en uva y ésteres quínicos en otras frutas.
- *Ácidos benzoicos* (ej. ácido gálico): Se encuentran normalmente en forma de benzoatos. La hidrólisis de éstos produce los ácidos libres en vino.
- *Estilbenos* (ej. resveratrol): Estilbenos hidroxidados se encuentran en las uvas y otras fuentes alimenticias. Los produce la vid en todos los tejidos como fitoalexinas en respuesta al ataque de hongos, como *Botrytis cinérea*.

## 1.2. Maduración

Antes de abarcar este asunto es necesario hacer una reflexión sobre el significado de la madurez de la semilla de uva. Aunque se conozcan en profundidad los cambios químicos que sufren, no existen etapas bien delimitadas que permitan hablar de estados de maduración de semillas. Es bien conocido que semillas de uva insuficientemente maduras aportan componentes indeseables que confieren al vino un carácter agresivo. Al contrario, también se sabe que semillas maduras dan al vino características de astringencia, estabilidad, cuerpo y aroma que estarán en armonía con el resto de componentes. Pero estas dos situaciones extremas comprenden muchos estados intermedios. La no existencia de directrices generales para la definición de la madurez sucede por una parte porque son muchas las características las que cambian en la semilla durante ese periodo. No sólo en composición cualitativa y cuantitativa, sino también por cambios en la humedad de la semilla y en características que alterarán la extractabilidad de estos compuestos durante la vinificación, como son textura, dureza, lignificación, etc. Por otro lado, la diversificación del mercado y los cambios en las preferencias de los consumidores hacen que las uvas vendimiabiles puedan tener semillas en

diferentes estados de madurez según sea el propósito en la vinificación. Así, el estado de las semillas para elaborar un vino tinto con intención de someterlo a un envejecimiento en barrica será diferente del estado de las semillas para la elaboración de un vino blanco joven con sabor y aroma herbáceos. Sabiendo que los fenómenos de polimerización, lignificación, oxidación, endurecimiento, desecación, etc. de las semillas suceden al mismo tiempo, en la bibliografía se suele hablar sin distinción de madurez fenólica, madurez tánica, o simplemente madurez de las semillas.

Según Ristic *et al.* (2005), *“madurez de semillas y madurez tánica son términos usados indistintamente en viticultura y enología con un amplio intervalo de definiciones para cada uno de los ellos”*. Por ejemplo, se puede definir la madurez de la semilla como *“un estado donde tanto la acumulación de reservas como el peso seco de la semilla han alcanzado un máximo”* (Austin, 1972). No obstante, ese estado se consigue poco después del envero, y la semilla en este punto no es madura para otros autores. Una vez que se ha completado este proceso de acumulación de reservas, la capa más externa se va volviendo oscura, dura e impermeable. Al final de estos procesos es cuando se alcanza la madurez de las semillas para Kennedy *et al.* (2000a), cuando la extractabilidad de los compuestos tánicos es mínima y sus características sensoriales son las menos agresivas posibles.

Por esta falta de unanimidad en la definición de madurez, se hablará en adelante de “madurez fenólica” o “madurez de la semilla” al estado en el cual la semilla es idónea para ser vendimiada sin referirse a un estado químico concreto, y “maduración” a todo el proceso por el cual las semillas se van desarrollando hasta llegar a ese estado óptimo.

### 1.2.1. Aspectos generales y desarrollo de la uva

El desarrollo de las semillas de uva desde la floración hasta la vendimia se puede englobar dentro del desarrollo del fruto entero. Éste a su vez se puede



---

dividir en tres etapas (Figura 5): (a) La primera fase comienza con la formación del fruto y está caracterizada por un aumento continuo de la semilla y la baya. Es en este período cuando se biosintetizan y se acumulan los compuestos fenólicos en las semillas, que tienen un color verde predominante. (b) Le sucede una etapa latente en la que se frena el crecimiento de la baya. Esta etapa es fácilmente reconocible, pues coincide con el envero de la uva, cuando se produce la pérdida de clorofilas y comienza la síntesis de los colorantes definitivos en el hollejo. (c) En la última etapa, la baya sigue creciendo y los ácidos de la pulpa se transforman en azúcares. Es durante esta etapa cuando se producen los cambios químicos de mayor interés en las semillas de uva.

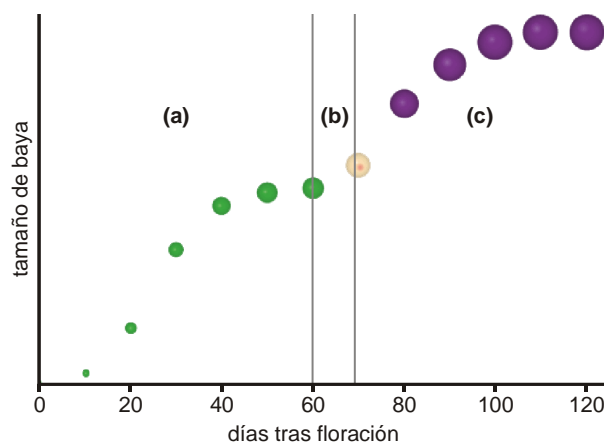


Figura 5. Esquema general de la evolución de la uva

### 1.2.2. Problemática de los climas cálidos

En climas cálidos, el cultivo de la vid y la elaboración de vinos de calidad se enfrentan a problemas relacionados con las altas temperaturas ambientales. Con primaveras secas y veranos calurosos, el periodo comprendido entre el envero y la madurez industrial de la uva disminuye, mientras que aumenta el desfase entre la madurez de la pulpa y las partes sólidas. Esto dificulta que las uvas alcancen la correcta madurez aromática y fenólica, debido a una reducción de los reguladores endógenos de la acumulación de antocianos en los hollejos, produciendo colores pobres o irregulares (Bergqvist et al., 2001; Zamora-Marín, 2006). Estos problemas son bien conocidos por los enólogos, de manera que las “caídas de color” en vinos

producidos en climas cálidos han llevado en muchas regiones a desestimar las elaboraciones de tintos, o a considerar la aplicación de técnicas más o menos avanzadas para promover la síntesis de pigmentos o su extracción durante la maceración como es la maceración prefermentativa en frío (Gómez-Míguez et al., 2007; Zamora-Marín, 2004). Asimismo, otras técnicas se aplican para favorecer la estabilización del color durante la vinificación. En este sentido, el contenido fenólico procedente de las semillas de la uva juega un papel importante por su implicación en la copigmentación de los antocianos. El estado de maduración de la uva, y en concreto la madurez fenólica de las semillas, resultan de extraordinaria importancia en la producción de vinos tintos de calidad en las zonas más afectadas por los efectos del cambio climático.

### 1.2.3. Desarrollo fisiológico de la semilla de uva

Con pequeñas diferencias varietales, las semillas de uva tienen la forma que muestra la Figura 6. El pico inferior puede ser recto o curvo, y arrugado o liso. También varía en grosor y en el ángulo que forma en el hilio. En el lado opuesto al pico, existe una muesca con una hendidura más o menos pronunciada dependiendo de la variedad. La parte dorsal es gruesa y redondeada (Figura 6b), y tiene en medio una zona circular llamada “chalaza”, que a su vez puede ser abultada o hundida. La parte ventral (Figura 6a) se caracteriza principalmente por una marca (rafe), que comienza en el pico y rodea longitudinalmente la semilla rodeando la muesca hasta llegar a la chalaza. También en la zona ventral se pueden encontrar unos hoyuelos a ambos lados del rafe. En total, están compuestas de agua (25-45%), polisacáridos (34-36%), taninos (4-10%), compuestos nitrogenados (4-6.5%), minerales (2-4%) y lípidos (13-20%), aunque estas proporciones van cambiando durante su desarrollo.

Una semilla madura está formada por tres tipos de tejidos: una cubierta formada por los tegumentos externo e interno, un endosperma de contenido adiposo y un embrión (Figura 6c-d). La cubierta exterior está formada por muchas capas de células que contienen compuestos fenólicos (Thorngate & Singleton,

---

1994) y cristales de oxalato cálcico (Pratt, 1971). El endosperma es un tejido blando dentro del tegumento interno formado de manera uniforme por células poligonales que contienen aceites y granos de aleuronas con cristales de oxalato cálcico (Pratt, 1971).

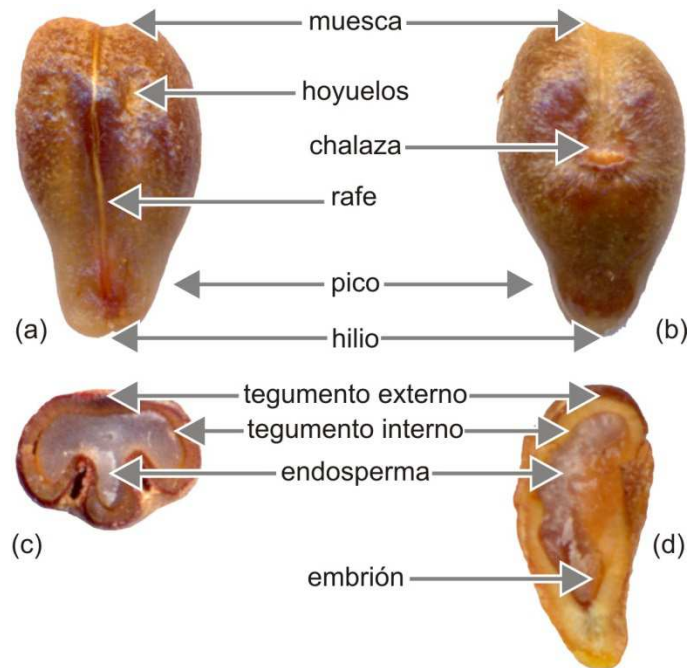


Figura 6. Vista ventral (a) y dorsal (b), y cortes transversal (c) y longitudinal (d) de una semilla de uva

Durante la vida de la semilla, cada una sufre tres procesos fisiológicos: (i) crecimiento, (ii) acumulación de nutrientes (grasas, proteínas y almidón) e (iii) inicio de la etapa latente. Cada proceso se caracteriza por cambios químicos y estructurales. En particular, la cubierta exterior es la que experimenta cambios más acentuados durante el desarrollo de la semilla.

### *CRECIMIENTO*

Cada baya tiene un total de cuatro óvulos dentro de los dos carpelos en el ovario. Aunque cada uva está fisiológicamente preparada para contener cuatro semillas, el número de ellas suele ser menor que cuatro. Éstas pueden distinguirse unas tres semanas después de la floración, de color verde blanquecino y muy blandas y flexibles.

### *ACUMULACIÓN DE NUTRIENTES*

Sobre el día 40 después de la floración la semilla ha alcanzado el 90% de su tamaño final, ha formado la chalaza y su color, tras haber pasado una etapa verdosa más intensa, comienza a amarillear. En la cubierta comienzan a sintetizarse monómeros de flavan-3-oles, la mayoría catequina, epicatequina y galoepicatequina (Thorngate & Singleton, 1994). La poca rigidez de los tegumentos permite que puedan actuar de transporte de nutrientes desde el pericarpio hasta el saco embrionario (Werker, 1997).

### *INICIO DE LA ETAPA LATENTE*

Dos meses tras la floración, las semillas han alcanzado su tamaño y peso fresco máximo. Los flavanoles dejan de sintetizarse, y éstos comienzan a evolucionar mediante reacciones de oxidación y polimerización a partir de uniones entre el carbono 4 y 8 de los monómeros simples (Figura 1). Los polímeros más sencillos, también llamados oligómeros, están compuestos por dos o tres monómeros; y polímeros de más de tres unidades básicas ya son considerados procianidinas o taninos. Al final de la maduración el contenido en monómeros y oligómeros es alrededor del 10% del total de flavanoles en las semillas de uva, el resto se han polimerizado u oxidado. Existen muchos estudios sobre los cambios químicos que sufren las semillas durante su maduración (Ferrer-Gallego et al., 2010; Kennedy et al., 2000a; Kennedy et al., 2000b; Liu et al., 2010; Peyrot des Gachons & Kennedy, 2003). A la vez que estos cambios suceden, se producen otros cambios físicoquímicos en la cubierta de las semillas, como la pérdida de agua y el endurecimiento de la semilla provocado por la lignificación de las capas más internas del tegumento externo. El endurecimiento surge a la semilla de protección contra la pérdida total de humedad y los daños mecánicos (Werker, 1997). También sirve de impermeabilidad al agua y al oxígeno exterior. Así se aumenta la capacidad de latencia y se aísla al embrión de procesos oxidativos. (Werker, 1980). La capa generada es gruesa en la parte del pico y más fina y débil en la zona de los hoyuelos en la parte ventral. Mientras, el tegumento interno permanece fino y se adhiere al endosperma firmemente (Mullins, Bouquet, and Williams, 1992). El

---

endurecimiento del tegumento externo viene acompañado de cambios importantes en el color, conociéndose este fenómeno como pardeamiento de la semilla.

Los cambios en el color de la semilla están relacionados con cambios en la composición fenólica, y existen estudios que demuestran una fuerte correlación entre el color y la cantidad de taninos extraíbles. Es por ello que el color de la semilla se considera tradicionalmente uno de los criterios más utilizados para evaluar la madurez de la semilla de uva (Ristic & Iland, 2005).



---

## 2. EL COLOR

El color es una respuesta mental al estímulo que produce una radiación luminosa visible en la retina. Se considera un concepto psicofísico debido a la contribución del fenómeno físico y de la sensación que provoca. Wyszecky y Stiles (1982) dan una definición más precisa:

*“El color es el aspecto de la percepción visual por el cual un observador puede distinguir diferencias entre dos campos de visión del mismo tamaño, forma y estructura, causadas por diferencias en la composición espectral de la energía radiante implicada en la observación”.*

Por tanto, no es una propiedad intrínseca del objeto, pues basta cambiar la fuente de iluminación para modificar el estímulo producido. Estímulo que depende de la distribución espectral de la radiación incidente, la capacidad del objeto para interactuar con esa radiación y de la fisiología del observador.

El fenómeno del color ha sido estudiado desde la antigüedad. Ya en la Grecia Clásica, Platón contemplaba la idea del color como un fuego interior que da origen a unos rayos visuales que se dirigen desde el ojo hacia el objeto. Epicuro fue más acertado al proponer que era el ojo el que recibía emanaciones provenientes del objeto. Sin embargo, al desconocerse la teoría ondulatoria, justificaba el fenómeno mediante la existencia de diminutos corpúsculos coloreados. Aristóteles fue el primero que vaticinó la relación entre los fenómenos de la luz y el color en su obra

*De Coloribus*, y argumentó que eran los objetos los que conferían “negrura” a la luz blanca resultando a partir de ahí los colores.

Hubo que esperar hasta el s. XVII para que Isaac Newton (1671) asentara muchas de las bases de la colorimetría que perduran hoy día. Newton describe el color como un fenómeno puramente físico y lo comprobó en sus experiencias al descomponer la luz solar en los siete colores espectrales al pasar por un prisma. En sus cartas a la Royal Society explicaba que el color era una magnitud tridimensional. Las tres sensaciones que dan al color este carácter son:

- *Matiz o tono*: Es el atributo fundamental y se define tradicionalmente como el que hace clasificar los colores como rojizos, verdosos, etc. Está relacionado con las diferencias de absorbanza de la energía radiante a diferentes longitudes de onda ( $\lambda$ ) en el espectro visible. Es el atributo cualitativo del color.
- *Brillo o luminosidad*: Es la característica por la cual un color puede clasificarse como más claro o más oscuro. Este atributo cuantitativo no es más que una medida relativa de la cantidad total de luz reflejada o transmitida frente a la cantidad de luz incidente.
- *Saturación o pureza*. Es también un atributo cuantitativo, y determina para cada tono su grado de diferencia respecto al color gris de la misma luminosidad.

Más adelante, Goethe (1749-1832) aborda el problema desde un punto de vista fisiológico, y difiere con Newton afirmando que el color depende del individuo y es subjetivo. Aquí comienza la dualidad entre la componente física y psíquica del color. La teoría psicofísica adquirió más importancia con los estudios de George Palmer en 1777 sobre la fisiología de la retina y los estudios en 1794 de Dalton sobre las anomalías cromáticas.

En el s. XIX aparece otra figura importante, Thomas Young (1773-1829), médico, físico y egiptólogo, quien predice la Teoría Tricromática del Color. Explica



---

que en la retina existen tres detectores diferentes de cuya conjugación se consigue predecir el estímulo de color (Young, 1802a; Young, 1802b):

*“Es imposible concebir que cada punto sensitivo de la retina esté compuesto por un número infinito de partículas capaces de vibrar en perfecto unísono con cada posible ondulación, por lo que hay que suponer un número limitado de las mismas sensible a los tres colores principales, capaces de ponerse en movimiento en mayor o menor concordancia según las diferentes ondulaciones. Cada filamento sensitivo del nervio óptico consiste en tres porciones, una para cada color”.*

Herman Ludwig von Helmholtz contribuye con sus estudios en el campo de la visión; y Grassmann y Maxwell completan la Teoría Tricromática con sus estudios sobre mezclas de colores. Dedujeron que las luces del mismo color, pero de diferente composición espectral, se comportan de forma idéntica en las mezclas, lo que condujo a la consideración de las Ecuaciones de Igualación de Color. *“Para igualar cualquier color basta con la determinación de las cantidades de tres luces primarias estandarizadas” (Maxwell, 1860).*

Hasta entonces, las imprecisiones de esta teoría eran consecuencia de las limitaciones tecnológicas. No existían fuentes de luz estables, no era posible producir luz monocromática de alta intensidad ni había medios para medir la luz de manera fiable. Los avances en tecnología, biología y psicología a principios del s. XX permitieron mejorar y ampliar estas teorías. Merece la pena nombrar los trabajos de König sobre las anomalías en la visión del color y su teoría de los colores oponentes, que suponía la existencia de un mecanismo de colores antagónicos, y que se diferenciaba en tres procesos visuales. Teoría que fue completada por Adams (1923) y Judd (1949):

$$R=-G \quad B=-Y \quad W=-N$$

Actualmente, todas las teorías y los sistemas de color están regularizados por la Comisión Internacional de Iluminación, CIE (*Commission Internationale de l'Eclairage*). Este organismo es el encargado de compendiar, evaluar y recomendar los avances en Colorimetría.

## 2.1. Espacio triestímulo

Las leyes que rigen la generalización tricromática se pueden expresar de forma geométrica, considerando un espacio de tres dimensiones denominado Espacio Triestímulo, en el que cada estímulo de color (Q) viene expresado por un vector tridimensional, donde cada dimensión es proporcional a cada uno de los estímulos primarios rojo (R), verde (G) y azul (B).

Para especificar completamente el estímulo de color es necesario además conocer la distribución espectral de la energía radiante con la cual se ilumina una superficie y el tamaño angular de dicha superficie. Por ello, y para simplificar el desarrollo de las expresiones para la igualación de color, estos factores se suelen considerar fijos. Con lo cual, el estímulo solamente viene caracterizado por su distribución espectral de energía,  $\{P_\lambda d_\lambda\}$ . Así, se definen los estímulos primarios de valor unitario:

$\{\rho P_\lambda d_\lambda\}R$ , que define el estímulo primario ROJO

$\{\gamma P_\lambda d_\lambda\}G$ , que define el estímulo primario VERDE

$\{\beta P_\lambda d_\lambda\}B$ , que define el estímulo primario AZUL

$\{P_\lambda d_\lambda\}Q$ , para un estímulo de color cualquiera, Q

$\rho$ ,  $\gamma$ , y  $\beta$  son factores constantes positivos que tienen utilidad en determinadas aplicaciones. En general se suelen escoger como cantidades unitarias de estímulos primarios aquellas que al mezclar aditivamente producen una igualación de color completa con un estímulo acromático especificado, es decir, con un espectro de energía constante.

Según esto, cualquier estímulo Q podría expresarse como la mezcla aditiva de determinadas cantidades de los tres estímulos primarios fijados R, G y B.

$$Q \equiv R_Q R + G_Q G + B_Q B \quad \text{Ec. 1}$$

Los factores escalares  $R_Q$ ,  $G_Q$ ,  $B_Q$  asignados a los estímulos primarios se denominan **valores triestímulos** (término abreviado de “Valores del Triple

Estímulo Internacional”) de  $Q$ . Los vectores unitarios  $R$ ,  $G$ ,  $B$ , que representan los estímulos primarios, definen el espacio vectorial, tienen un origen común y se sitúan en tres direcciones arbitrariamente elegidas. Así, en el espacio triestímulo- $(R,G,B)$ , el estímulo  $Q$  se define como un vector triestímulo- $(R_Q,G_Q,B_Q)$ .

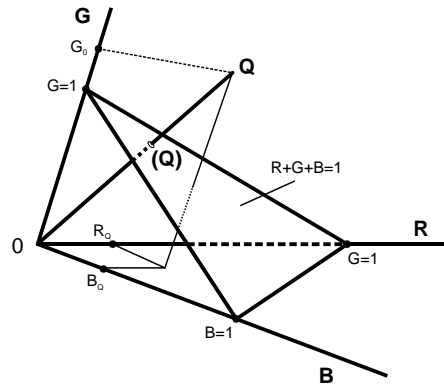


Figura 7. Espacio cromático  $(R,G,B)$

Una representación bidimensional útil es la que se obtiene al proyectar el vector triestímulo  $Q$  sobre el plano  $R+G+B=1$  (Figura 7). Esta proyección define la dirección del vector, pero no la intensidad del mismo. Por convenio, en ese plano se eligen como ejes de coordenadas las líneas de intersección entre dicho plano y cada uno de los planos definidos por pares de vectores primarios. El triángulo resultante se denomina *diagrama de cromaticidad*- $(r,g)$ . Las coordenadas de este espacio están relacionadas con los valores triestímulo y vienen definidas por las ecuaciones:

$$r = \frac{R}{R + G + B} \quad g = \frac{G}{R + G + B} \quad b = \frac{B}{R + G + B} \quad \text{Ec. 2}$$

De este modo, es suficiente conocer dos de las coordenadas para definir los valores triestímulo, y se simplifica su representación bidimensional. El triángulo equilátero formado se conoce como *Triángulo de Color de Maxwell* (Figura 8). Se suele representar sobre un triángulo donde los ejes de coordenadas elegidos  $(r,g)$  sean perpendiculares. Esta modificación en la representación no altera ninguna de las ecuaciones colorimétricas básicas.

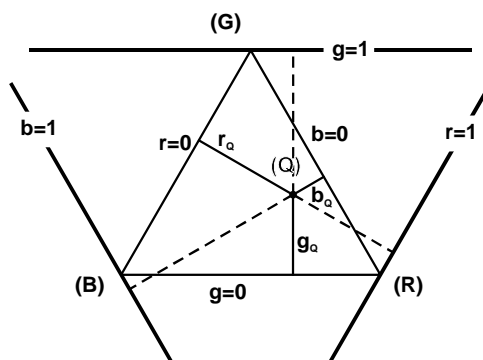


Figura 8. Triángulo de Maxwell

## 2.2. Ecuaciones colorimétricas básicas

En un espacio vectorial, un vector puede representarse como las sumas parciales de sus componentes. También un estímulo de color  $Q$ , formado por una distribución de energía radiante  $\{P_\lambda d\lambda\}_Q$ , puede expresarse como la mezcla aditiva de una serie de estímulos  $Q_i$ , con distribuciones espectrales  $\{P_\lambda d\lambda\}_{Q_i}$ , asociado a  $n$  intervalos de longitud de onda de ancho  $(\Delta\lambda)_i$ . (donde  $i=1$  a  $n$ ). Los intervalos, que pueden tener anchura variable, dividen el espectro visible (380–770 nm).

Cada estímulo  $Q_i$  se define por la ecuación básica:

$$Q \equiv R_{Q_i} R + G_{Q_i} G + B_{Q_i} B \quad \text{Ec. 3}$$

donde  $R_{Q_i}$ ,  $G_{Q_i}$  y  $B_{Q_i}$  son los valores triestímulo de  $Q_i$ . La mezcla aditiva de todos los estímulos  $Q_i$  igualará el estímulo original  $Q$ , pudiendo definirse como:

$$Q = \sum_{i=1}^n Q_i = \left( \sum_{i=1}^n R_{Q_i} \right) R + \left( \sum_{i=1}^n G_{Q_i} \right) G + \left( \sum_{i=1}^n B_{Q_i} \right) B \quad \text{Ec. 4}$$

El número de intervalos  $n$  aumenta conforme se reduce la anchura de los mismos, y en cuyo caso límite resulta:

$$P = \lim_{|(\Delta\lambda)_{i=k}| \rightarrow 0} \sum_{i=1}^n P_{\lambda_i} (\Delta\lambda)_i \quad \text{Ec. 5}$$

---

esto es, la integral de la energía radiante espectral  $P_\lambda$  entre los dos extremos del espectro visible ( $\lambda_{380}$ - $\lambda_{770}$ ) y donde se consideran normalmente intervalos constantes de longitud de onda:

$$\int_{\lambda_{380}}^{\lambda_{770}} P_\lambda d\lambda = \lim_{\Delta\lambda \rightarrow 0} \sum_{i=1}^n P_{\lambda_i} (\Delta\lambda)_i \quad \text{Ec. 6}$$

donde  $\lambda_i$  corresponde a la longitud de onda central del intervalo  $\Delta\lambda$ ,  $P_\lambda \Delta\lambda$  representa la energía radiante del intervalo  $d\lambda$ , y se denomina normalmente “estímulo monocromático de longitud de onda  $\lambda$ ,  $Q_\lambda$ ”. Cada estímulo monocromático  $Q_\lambda$  puede expresarse como:

$$Q_\lambda \equiv R_\lambda R + G_\lambda G + B_\lambda B \quad \text{Ec. 7}$$

donde  $R_\lambda$ ,  $G_\lambda$  y  $B_\lambda$  se denominan “valores triestímulos espectrales”.

### 2.3. Sistema colorimétrico CIE

A partir de las bases anteriores, la CIE propuso en 1931 su sistema colorimétrico basado en unas “Especificaciones sobre observadores estándares en colorimetría” y unas “Recomendaciones sobre cuestiones relacionadas” (CIE, 1932).

La dificultad en el cálculo de los valores triestímulos R, G, B según las ecuaciones básicas indujo a la CIE a la adopción de otro sistema tricromático. Mediante una transformación matricial, a partir de los primarios R, G, B se definen los nuevos estímulos X, Y, Z. Además, se definen las “funciones de igualación de color” como las propiedades del Observador Estándar CIE-1931 (Observador de 2°), en el intervalo de longitudes de onda de la zona visible del espectro. Se representan por:

$$\bar{x}(\lambda) \quad \bar{y}(\lambda) \quad \bar{z}(\lambda)$$

En 1964 se define un Observador Estándar Colorimétrico Suplementario (u Observador de 10°), análogo al anterior, cuyas funciones de igualación son:

$$\bar{x}_{10}(\lambda) \quad \bar{y}_{10}(\lambda) \quad \bar{z}_{10}(\lambda)$$

Una vez definidos los observadores estándares, la CIE define una serie de distribuciones espectrales correspondientes a los “Iluminantes Estándares CIE”, simbolizados por la expresión  $\{S(\lambda) d(\lambda)\}$ , donde  $S(\lambda)$  representa la concentración espectral de energía radiante incidente (Judd & Wyszecki, 1975; Wyszecki & Stiles, 1982).

Entre los diferentes iluminantes propuestos están:

- *Iluminante A*: Representa la luz de un radiador total a la temperatura absoluta de 2856 K, según la Escala Internacional Práctica de Temperatura 1968.
- *Iluminante B*: Representa la luz solar directa y se relaciona con una temperatura de color aproximada de 4874 K.
- *Iluminante C*: Se asimila con la luz diurna media (tradicionalmente luz de día nublado), con temperatura de color de 6774 K aproximadamente. Este iluminante ha sido ampliamente usado en la evaluación del color en alimentos, aunque en 1971 fue eliminado junto con el iluminante B, de la lista de iluminantes recomendados por presentar serias deficiencias en la distribución espectral en la región UV, importante en el caso de materiales fluorescentes.
- *Iluminante D65*: Se relaciona con una temperatura de color aproximada de 6504 K y representa la luz de día natural. Es el más utilizado y el que la CIE recomienda en la mayoría de los casos.
- *Iluminante D55*: Relacionado con una temperatura de color de 5503 K.
- *Iluminante D75*: Éste y el D55 son la alternativa al D65 recomendada por la CIE. Se relaciona con una temperatura de color de 7504 K.

---

## 2.4. Sistemas CIE 1931(x,y) y CIE 1964-(x10,y10) (CIEXYZ)

Para la definición de los valores triestímulos y el cálculo de las coordenadas de cromaticidad, la CIE propone ecuaciones generales, donde cada estímulo surge de la integración de las componentes monocromáticas de todas las longitudes de onda a lo largo del espectro visible teniendo en cuenta las funciones de igualación según el observador seleccionado. De ahí, se obtienen los valores triestímulos X,Y,Z, y al normalizar éstos, las coordenadas de cromaticidad x,y,z.

$$x = \frac{X}{X+Y+Z} \quad y = \frac{Y}{X+Y+Z} \quad z = \frac{Z}{X+Y+Z} \quad \text{Ec. 8}$$

De esta manera, conociendo las coordenadas de cromaticidad x e y, z queda definida como la diferencia de x e y hasta uno. Este cálculo simplifica la representación bidimensional del color. Así, el diagrama cromático CIE 1931-(x,y) surge de la representación en el diagrama x-y de los valores triestímulos de los estímulos monocromáticos (Figura 9). El perímetro formado por los estímulos correspondientes a las longitudes de onda puras se denomina "spectrum locus", y la línea recta de púrpuras representa las diferentes mezclas entre los estímulos monocromáticos extremos. Merece la pena remarcar que este diagrama no es coloreado. Sin embargo, se ha rellenado en la figura para ilustrar la distribución de estímulos cromáticos en el espacio de color.

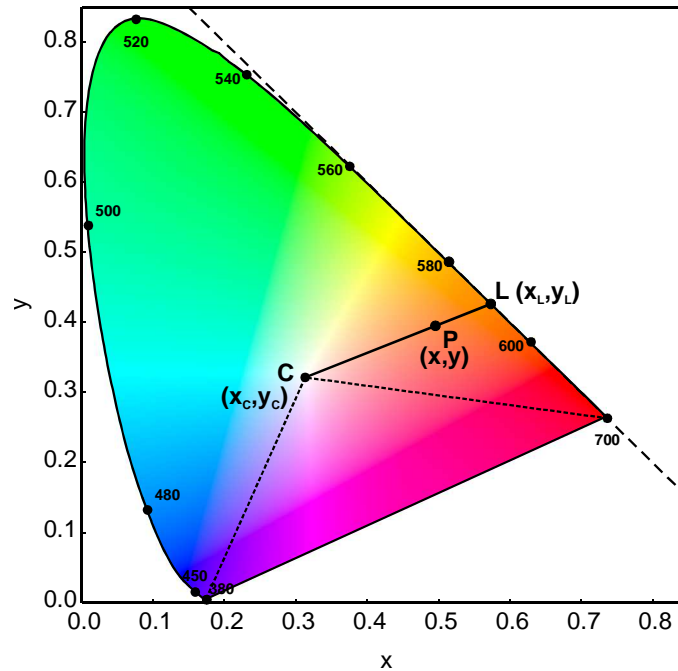


Figura 9. Diagrama cromático CIE 1931-(x,y)

Este espacio de color introdujo dos conceptos importantes en la caracterización del color: la longitud de onda dominante ( $\lambda_d$ ) y la pureza o saturación:

- La longitud de onda dominante de un estímulo coloreado es la correspondiente al estímulo monocromático que, mezclado con uno acromático (iluminante de referencia) produce la igualación con el color original. Si el estímulo coincide en la “línea de los púrpuras” (zona no espectral del “spectrum locus”), se utiliza el concepto de longitud de onda complementaria ( $\lambda_c$ ), y corresponde a la del estímulo monocromático espectral linealmente opuesto.
- La saturación o pureza se define como el porcentaje de estímulo monocromático de  $\lambda_d$  en la mezcla que iguala el estímulo considerado. Se calcula mediante la relación de distancias al punto de acromaticidad (iluminante). Considerando el punto  $P$  de la figura, su pureza de color vendría dada por la siguiente expresión:



---

$$Pureza(\%) = \frac{\overline{CP}}{\overline{CL}} \times 100$$

Ec. 9

El Espacio Cromático CIE 1931 permite cuantificar el color. Sin embargo, algunos trabajos pusieron de manifiesto la falta de uniformidad de este espacio. David L. MacAdam (1942) estudió regiones del espacio de color CIE 1931-(x,y) cuya apariencia era la misma para un observador estándar. De acuerdo con la forma en la que estas regiones se distribuían, ajustó estas regiones a elipses cuyas dimensiones estaban expresadas en términos de desviaciones estándar dentro de los espacios de color.

En búsqueda de un espacio de color uniforme, se desarrolló la Escala Uniforme UCS 1960, y los espacios U\*V\*W y CIELUV, aunque no fue hasta 1976 cuando la CIE propuso el espacio de color CIELAB, que es sin duda el más usado en colorimetría.

## 2.5. Espacio CIELAB

El espacio de Color CIE 1976 L\*a\*b\* (CIELAB), que procede de los sistemas de *raíz cúbica* (Glasser et al., 1958) como el sistema ANLAB (Glasser & Troy, 1952) queda definido, tal como muestra la (Figura 10), cuando se representan en coordenadas rectangulares las siguientes expresiones para claridad L\*, y cromaticidad, a\* y b\*. Al igual que en la figura anterior, el color solo tiene fines ilustrativos.

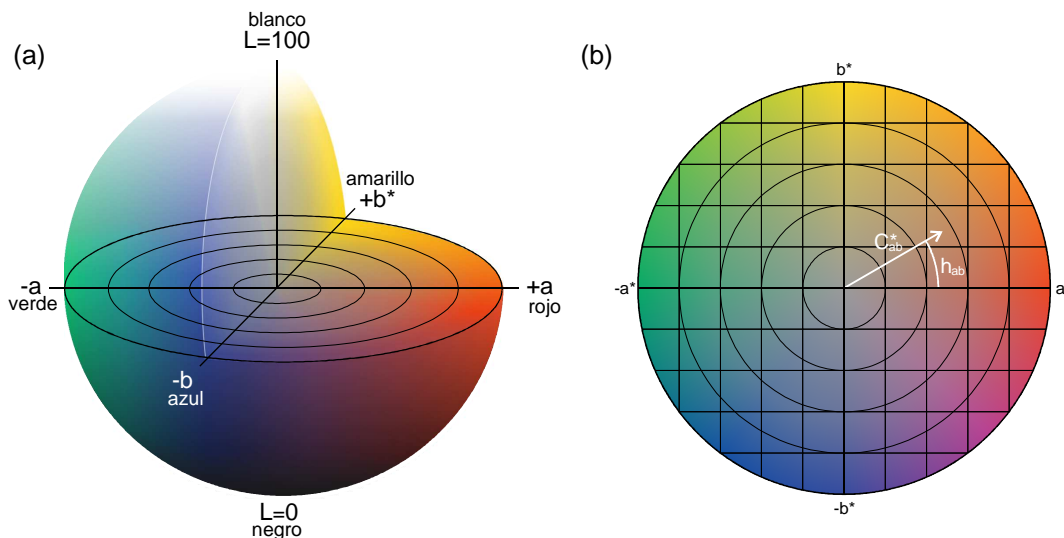


Figura 10. (a) Espacio de Color CIE 1976- $(L^*a^*b^*)$ . (b) Vista del plano  $(a^*,b^*)$  mostrando las coordenadas polares  $C_{ab}^*$  y  $h_{ab}$

Las variables colorimétricas dependerán entonces de los valores triestímulo  $X, Y, Z$  y de  $X_n, Y_n, Z_n$ . Además, y para corregir pérdida de uniformidad, Pauli (1976) introdujo ciertas modificaciones para  $X/X_n, Y/Y_n$  o  $Z/Z_n < (24/116)^3$ .

— *Claridad* ( $L^*$ ): Es la magnitud por la que un color se clasifica como más claro o más oscuro. Sus límites son 0 y 100 y corresponden al negro y blanco puros respectivamente:

$$L^* = 116 \sqrt[3]{\frac{Y}{Y_n}} - 16 \quad \text{para } Y/Y_n > (24/116)^3 \quad \text{Ec. 10}$$

$$L^* = 9033 \frac{Y}{Y_n} \quad \text{para } Y/Y_n \leq (24/116)^3 \quad \text{Ec. 11}$$

— *Cromaticidad* ( $a^*, b^*$ ):

$$a^* = 500 \left[ f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \quad \text{Ec. 12}$$

$$b^* = 200 \left[ f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \quad \text{Ec. 13}$$

---

siendo:

$$f\left(\frac{X}{X_n}\right) = \sqrt[3]{\frac{X}{X_n}} \quad \text{para } X/X_n > (24/116)^3 \quad \text{Ec. 14}$$

$$f\left(\frac{X}{X_n}\right) = 7.787\left(\frac{X}{X_n}\right) + \frac{16}{116} \quad \text{para } X/X_n \leq (24/116)^3 \quad \text{Ec. 15}$$

$$f\left(\frac{Y}{Y_n}\right) = \sqrt[3]{\frac{Y}{Y_n}} \quad \text{para } Y/Y_n > (24/116)^3 \quad \text{Ec. 16}$$

$$f\left(\frac{Y}{Y_n}\right) = 7.787\left(\frac{Y}{Y_n}\right) + \frac{16}{116} \quad \text{para } Y/Y_n \leq (24/116)^3 \quad \text{Ec. 17}$$

$$f\left(\frac{Z}{Z_n}\right) = \sqrt[3]{\frac{Z}{Z_n}} \quad \text{para } Z/Z_n > (24/116)^3 \quad \text{Ec. 18}$$

$$f\left(\frac{Z}{Z_n}\right) = 7.787\left(\frac{Z}{Z_n}\right) + \frac{16}{116} \quad \text{para } Z/Z_n \leq (24/116)^3 \quad \text{Ec. 19}$$

Los valores triestímulo  $X_n$ ,  $Y_n$ ,  $Z_n$ . Por ejemplo, para el iluminante D65 y el Observador Estándar de 10°:

$$X_n=94.811 \quad Y_n=100 \quad Z_n=107.304$$

Al poderse considerar este espacio uniforme, la diferencia de color ( $\Delta E_{ab}^*$ ) queda definida como la distancia euclídea dentro de este espacio tridimensional:

$$\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad \text{Ec. 20}$$

Al convertir los valores de cromaticidad  $a^*$  y  $b^*$  en sus componentes polares, quedan definidos dos componentes del color, importantes en su especificación psicofísica:

— *Croma* ( $C_{ab}^*$ ):

$$C_{ab}^* = \sqrt{(a^*)^2 + (b^*)^2} \quad \text{Ec. 21}$$

— *Tono* ( $h_{ab}$ ):

$$h_{ab} = \arctan \left( \frac{b^*}{a^*} \right) \quad \text{Ec. 22}$$

que según los signos de  $a^*$  y  $b^*$  se definirá en cada cuadrante:

$$\begin{array}{llll} \text{si } a^* > 0, b^* > 0 & \rightarrow & 0^\circ < h_{ab} < 90^\circ \\ \text{si } a^* < 0, b^* > 0 & \rightarrow & 90^\circ < h_{ab} < 180^\circ \\ \text{si } a^* < 0, b^* < 0 & \rightarrow & 180^\circ < h_{ab} < 270^\circ \\ \text{si } a^* > 0, b^* < 0 & \rightarrow & 270^\circ < h_{ab} < 360^\circ \end{array}$$

Así, el croma queda definido por la distancia entre un punto del diagrama y el valor acromático de la misma claridad. Análogamente, el tono queda definido como un ángulo en grados sexagesimales, formado entre el vector que representa al punto en el diagrama y el eje de abscisas. En estas correlaciones psicométricas, de los conceptos de luminosidad, croma y matiz, con la percepción visual, hemos de tener en cuenta la influencia que pueden ejercer las condiciones de observación.

Para los cálculos de diferencias de tono en función de las diferencias de color, se utiliza la variable  $\Delta H_{ab}^*$ :

$$\Delta H_{ab}^* = \sqrt{(\Delta E_{ab}^*)^2 - (\Delta L^*)^2 - (\Delta C_{ab}^*)^2} \quad \text{Ec. 23}$$

Con objeto de mejorar la uniformidad del espacio CIELAB en cuanto a diferencias de color se refiere, la CIE ha actualizado estas fórmulas a CIE94 y CIEDE2000, donde se introducen algunos factores y constantes que mejoran la relación entre las diferencias de color calculadas y las diferencias percibidas (Melgosa et al., 2001).

---

---

### 3. VISIÓN ARTIFICIAL

La ciencia informática avanza hacia la creación de ordenadores cada vez más rápidos, expertos y autónomos. No sólo ahora, sino desde el primer ordenador, el hombre ha soñado con la creación de máquinas capaces de relacionarse con su entorno de la misma manera que lo hace un ser humano, con dispositivos que interactúen con el exterior tal y como lo hacen los sentidos. A pesar de las dificultades que conlleva, existe un especial interés en dotar a los ordenadores de uno de los cinco sentidos del hombre: la capacidad de ver. Las innumerables aplicaciones y los avances en sistemas ópticos han hecho que la visión artificial sea una de las técnicas con mayor campo de aplicación en la industria alimentaria.

De acuerdo con esto, la visión artificial se podría definir de una manera simple como *una disciplina que engloba todos los elementos y procesos que proporcionan ojos a una máquina* (González Marcos et al., 2006). Una definición más específica la da Nalwa (1993), diciendo que *la visión artificial o comprensión de imágenes describe la deducción automática de la estructura y propiedades de un mundo tridimensional, posiblemente dinámico, bien a partir de una o varias imágenes bidimensionales de ese mundo*. Nalwa además segrega las propiedades del mundo tridimensional en dos grupos: propiedades geométricas y propiedades materiales. Ejemplos de propiedades geométricas son el tamaño, la forma y la

posición de los objetos. Las propiedades materiales son la textura, la composición y el color, y serán abordadas con mayor profundidad en este trabajo.

Por otro lado, existen muchas analogías entre la visión de un ordenador y la visión de un ser humano. En los dos casos hay que adquirir la imagen e interpretarla. Para un ordenador, la cámara cumple las funciones fisiológicas del ojo, y es su sensor el que hace las veces de retina. Un ordenador procesando las imágenes realiza la función del cerebro, que debe reconocer, interpretar las formas y discernir qué información es útil en la imagen recibida. Pero un ordenador no es capaz de interpretar una imagen demasiado complicada. Para ello, haría falta una compleja estructura de procesadores interconectados a modo de neuronas, además de una cantidad extraordinaria de información preliminar de todos los objetos presentes en la imagen. Para evitar este tipo de problemas, se suelen utilizar escenas con una complejidad moderada. Así, para un ordenador será más fácil reconocer objetos que estén sobre un fondo plano o que proyecten una sombra definida.

Todo el proceso de análisis de imágenes abarca desde su captura hasta la interpretación de los resultados obtenidos. Da-Wen Sun (2000) divide las etapas secuencialmente según su grado de complejidad (Figura 11).

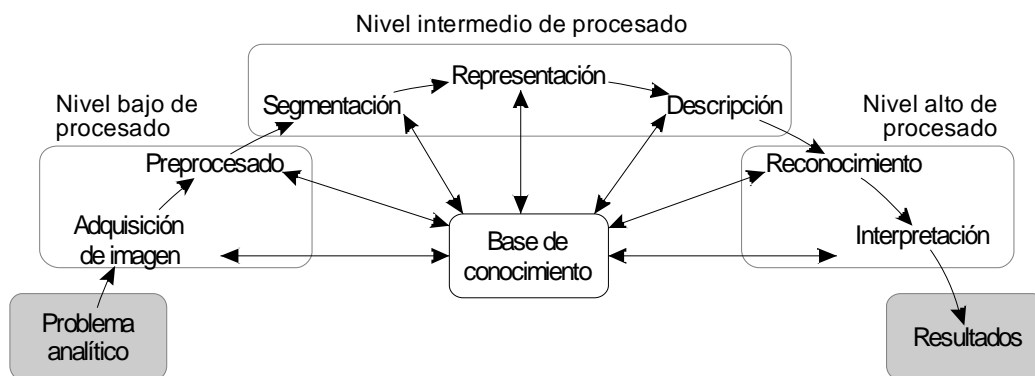


Figura 11. Diferentes niveles de procesado de imágenes (Sun, 2000).

Un nivel bajo de procesado incluye la adquisición de imagen y el preprocesado, es decir, el paso de la señal electrónica del dispositivo CCD a una

---

---

forma numérica y el tratamiento de la imagen cruda para corregir ruido, borrosidad, distorsiones provocadas por la óptica, etc... Los algoritmos que realizan estas tareas, aunque muchos de ellos son sencillos, consumen gran cantidad de tiempo. Esto es un inconveniente cuando se quiere procesar una escena en tiempo real. Para que un sistema de visión artificial funcione en tiempo real es necesario minimizar la carga computacional en la medida de lo posible. Las imágenes tratadas en este trabajo representan siempre escenas estáticas y no se necesitó una optimización tan exhaustiva del cálculo computacional en los algoritmos informáticos desarrollados. Además, para la medida del color, es de vital importancia las características de las imágenes sin preprocesar, tal y como son recogidas por la cámara. Por ello se ha evitado realizar tratamiento alguno entre la adquisición de la imagen y la segmentación.

Este proceso de segmentación es la etapa más importante del nivel intermedio de procesado. Su objetivo es dividir la imagen en regiones que pertenecen a una misma clase, y se abordará con más profundidad más adelante. En este nivel de procesado, la representación y descripción de la imagen se encargan de la extracción de información analítica de las partes seleccionadas tras la segmentación, donde se incluyen las medidas de color y morfología.

Por último, un nivel alto de procesado comprende el reconocimiento e interpretación de imágenes. El reconocimiento de imágenes es un proceso donde se etiquetan los objetos reconocidos mediante clasificación estadística o comparación con una base de datos; y por último la interpretación da sentido y valor a las características medidas en la imagen. Estas dos últimas etapas son el fin último de la visión artificial. Con este último nivel es posible evaluar el grado de maduración de un fruto, su estado de conservación o predecir la aparición de defectos y putrefacción en alimentos.

Todas las etapas del análisis de imagen requieren una base de datos que modele, enriquezca y perfeccione el modelo. Las particularidades de cada muestra modificará las condiciones en la adquisición de la imagen (nivel bajo), el proceso de segmentación (nivel intermedio) y los criterios para el reconocimiento y

clasificación (nivel alto de procesado). Por este motivo, la puesta a punto de la metodología es una tarea de crucial importancia para la obtención de resultados satisfactorios en el análisis de imagen.

### 3.1. Fundamentos de imágenes digitales

Según el diccionario de la lengua española (2001):

**imagen:**

(Del lat. *imāgo*, -*ñis*).

1. f. Figura, representación, semejanza y apariencia de algo.
3. f. *Ópt.* Reproducción de la figura de un objeto por la combinación de los rayos de luz que proceden de él.

**digital:**

(Del lat. *digitālis*).

2. adj. Referente a los números dígitos y en particular a los instrumentos de medida que la expresan con ellos.

A partir de las acepciones seleccionadas en cada caso se puede definir “imagen digital” como una representación visual de una figura compuesta por un conjunto numérico finito.

La naturaleza básica de una imagen, representada por  $f(x,y)$ , dependerá del objeto, la fuente radiante y el sistema de formación de la imagen. Una imagen digital puede ser representada por una matriz  $f$  de dimensiones  $N \times M$ :

$$f = \begin{bmatrix} f(1,1) & f(1,2) & \Lambda & f(1,M) \\ f(2,1) & f(2,2) & \Lambda & f(2,M) \\ M & M & f(i,j) & M \\ f(N,1) & f(N,2) & \Lambda & f(N,M) \end{bmatrix} \quad \text{Ec. 24}$$

donde cada elemento de la matriz, llamado píxel (en inglés, abreviatura de *picture element*), da la intensidad de la imagen en ese punto. El producto de  $M$  y  $N$  indica el



---

número de píxeles totales de la imagen. Por convenio y simplicidad, se suele usar el múltiplo megapíxel ( $10^6$  píxeles). Esta información sería suficiente para representar una imagen en escala de grises. De acuerdo con la teoría colorimétrica vista anteriormente, si se desea que cada píxel contenga información completa del color, son necesarias tres matrices como la descrita, una para cada estímulo de color, definiendo las dimensiones de la matriz  $f$  como  $N \times M \times 3$ .

### 3.1.1. Resolución

Cada píxel representa un área de la imagen real. En este aspecto, la resolución de una imagen indica el cociente entre la longitud de un píxel y la región del espacio que representa. La Figura 12 muestra el efecto de captar una misma imagen con distintas resoluciones. Puede observarse cómo la cantidad de información aumenta conforme lo hace la resolución de la imagen.

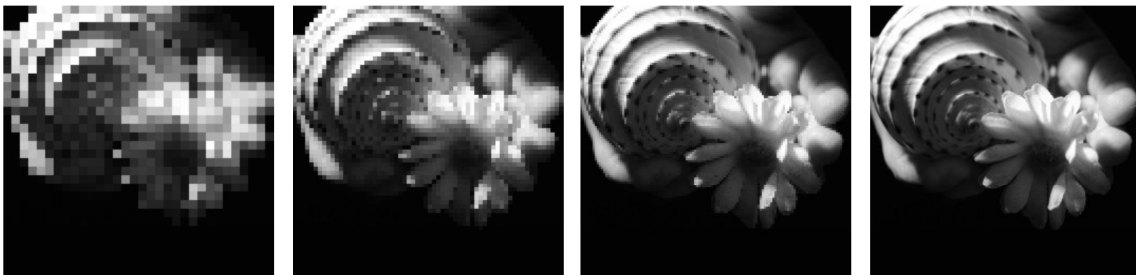


Figura 12. Imagen de  $32 \times 32$ ,  $64 \times 64$ ,  $128 \times 128$  y  $256 \times 256$  píxeles.

### 3.1.2. Profundidad de bits

El número de valores que puede tomar cada píxel no es infinito ni continuo. Al igual que todos los valores numéricos en informática, son almacenados en un sistema binario de numeración. Así, una profundidad de  $n$  bits implica que cada píxel de la imagen puede tener  $2^n$  valores ( $2^n$  colores distintos) para cada canal. La Figura 13 muestra la imagen anterior con distintas profundidades de bits. En la primera de las imágenes, de un bit, cada píxel puede tomar dos valores. La siguiente imagen, de dos bits, muestra gradaciones en cuatro niveles de gris. La última es una imagen común en escala de grises de 8 bits. Los 256 valores que

forman la paleta crean sensación de continuidad en la gradación del color, aunque esta continuidad no exista en realidad. Una cámara digital convencional almacena el color en cada píxel en gradaciones de los tres colores básicos con una profundidad de 8 bits por canal. Por ello, se habla normalmente de estas imágenes como imágenes en color de 24 bits.

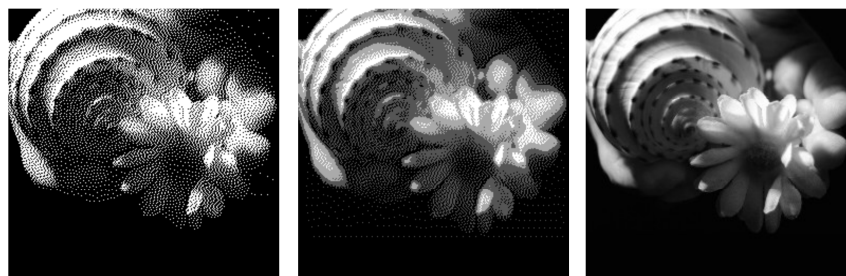


Figura 13. Efecto de la profundidad de bits en una imagen en blanco y negro.  
De izquierda a derecha: 1, 2 y 8 bits.

### 3.2. Adquisición de imágenes

Una cámara fotográfica digital es similar a una analógica en cuanto a óptica se refiere. La diferencia radica en el lugar donde se forman las imágenes. En la fotografía analógica la imagen se forma en la superficie de una película, la cual contiene unos finos cristales sensibles a la luz suspendidos en un gel; y es el tamaño de esos cristales el que define la sensibilidad de la fotografía.

En una cámara digital las imágenes se forman en la superficie de un dispositivo CCD (siglas en inglés de *charge-coupled device*: dispositivo de carga acoplada). Un CCD es un circuito que contiene un número de condensadores que son estimulados por la radiación registrando gradaciones de luz en los tres colores básicos rojo, verde y azul (RGB). El fundamento físico del fenómeno está basado en el efecto fotoeléctrico, mediante el cual, se produce una corriente eléctrica proporcional a la intensidad de luz recibida. Para separar los tres colores, los dispositivos CCD suelen contar con un Filtro Bayer (Figura 14). Éste está compuesto por un mosaico cuya unidad fundamental son cuatro píxeles, dos de ellos verdes, uno rojo y uno azul. Debido a esta disposición, la zona donde se encuentra un píxel rojo no hay sensibilidad a estímulos verdes y azules. Para

---

subsanoar esta carencia, se construye una se\u00f1al RGB a partir de los p\u00edxeles vecinos de cada color. Otros equipos m\u00e1s sofisticados resuelven este problema dividiendo la imagen en tres direcciones diferentes con la ayuda de prismas. Cada una de estas im\u00e1genes va a parar a un detector exclusivo para cada uno de los colores primarios, y la integraci\u00f3n final produce la imagen a color. Como desventaja, estos equipos necesitan m\u00e1s cantidad de luz, debido a la atenuaci\u00f3n provocada por la divisi\u00f3n de los haces.

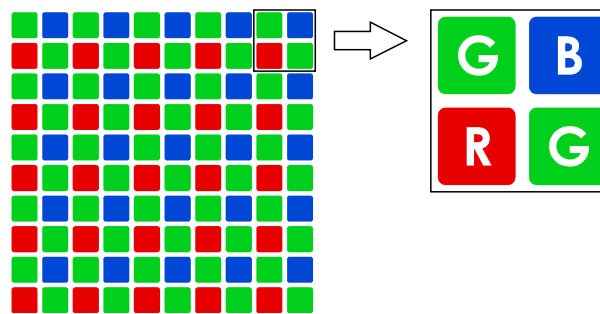


Figura 14. Disposici\u00f3n del filtro Bayer y su unidad elemental

Algunos equipos m\u00e1s modernos han sustituido los dispositivos CCD por sensores CMOS (siglas en ingl\u00e9s de *complementary metal oxide semiconductor*). Su principal ventaja es la velocidad. Estos dispositivos son capaces de funcionar a cientos o incluso miles de im\u00e1genes por segundo y ofrecen mayor linealidad de respuesta. Sin embargo, presentan algunas desventajas, como la existencia de un ruido denominado *fixed pattern noise*, adem\u00e1s de otros problemas que pueden ser corregidos mediante refrigeraci\u00f3n termoel\u00e9ctrica (Magnan, 2003).

### 3.3. Segmentaci\u00f3n

Como se vio anteriormente, segmentar una imagen digital consiste en dividirla en regiones con caracter\u00edsticas similares. Es decir, significa diferenciar los diversos objetos del fondo y localizarlos en el espacio. Al final de la etapa de segmentaci\u00f3n se tiene que poder extraer de cada objeto de la imagen sus caracter\u00edsticas individualmente (Gonz\u00e1lez Marcos et al., 2006). Asimismo, cada conjunto de p\u00edxeles pertenecientes a un mismo objeto puede tener una etiqueta

que lo defina, de forma que simplemente por agrupación de puntos con una misma etiqueta y conectados espacialmente, se pueda determinar la lista de objetos.

El resultado de este proceso queda reflejado en la máscara de segmentación. Una máscara de segmentación es una imagen bidimensional de las mismas dimensiones que la imagen a segmentar, donde mediante números enteros y discretos, cada píxel de la máscara indica si el píxel homólogo de la imagen original pertenece al fondo o a la *región de interés* (ROI, del inglés *region of interest*); y en su caso, con una etiqueta se muestra a qué objeto dentro de la región de interés se refiere. El caso más sencillo de una máscara de segmentación es una imagen binaria en el que el valor 1 se le asigna a la región de interés y el valor 0 se asigna al fondo de la imagen (Figura 15). En la imagen, las semillas no están etiquetadas, todas tienen el valor 1 en la máscara. Si la máscara de segmentación estuviera etiquetada, cada semilla tendría un número entero diferente. En este caso, habría píxeles con números del 1 al 4 según la semilla a la que pertenezcan.

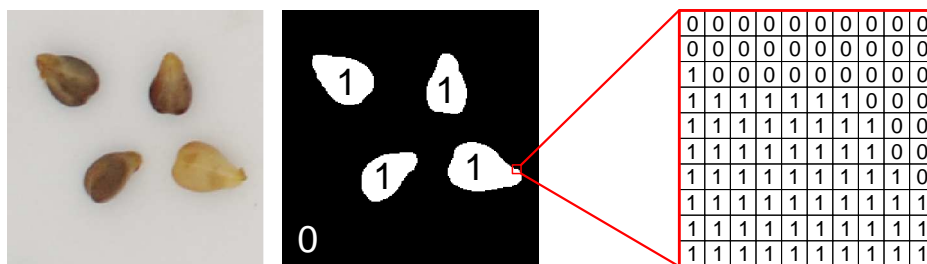


Figura 15. Máscara de segmentación de una imagen de semillas de uva mostrando las áreas cuyos píxeles tienen valores 1 y 0. A la derecha, detalle de los píxeles en una porción de esa máscara.

El método para realizar la segmentación dependerá de las características del fondo y de la región de interés. Además, existen transformaciones morfológicas que se pueden llevar a cabo para mejorar este proceso. Aunque existen más tipos de técnicas de segmentación y transformaciones morfológicas, se va a hacer hincapié sólo en las que se han aplicado en este trabajo.

---

### 3.3.1. Técnicas de segmentación

La *umbralización* es la técnica más utilizada para llevar a cabo la segmentación, y se basa en la distinta distribución de las frecuencias de intensidades (histograma) de los píxeles de la imagen. La Figura 16a muestra para una imagen en escala de grises cómo los píxeles de las semillas de uva tienen unas intensidades que quedan distribuidas por la parte media del histograma. Los píxeles pertenecientes al fondo de la imagen, mucho más numerosos, claros y homogéneos que los píxeles de las semillas, quedan agrupados en un pico estrecho y alto en la parte derecha. En la Figura 16b se puede ver cómo queda el aspecto de la máscara de segmentación según dónde se establezca el umbral.

Cuando la intensidad en escala de grises del fondo sea muy similar a la de la región de interés, se pueden utilizar los tres canales RGB, y establecer una condición diferente en cada uno de ellos para mejorar la segmentación.

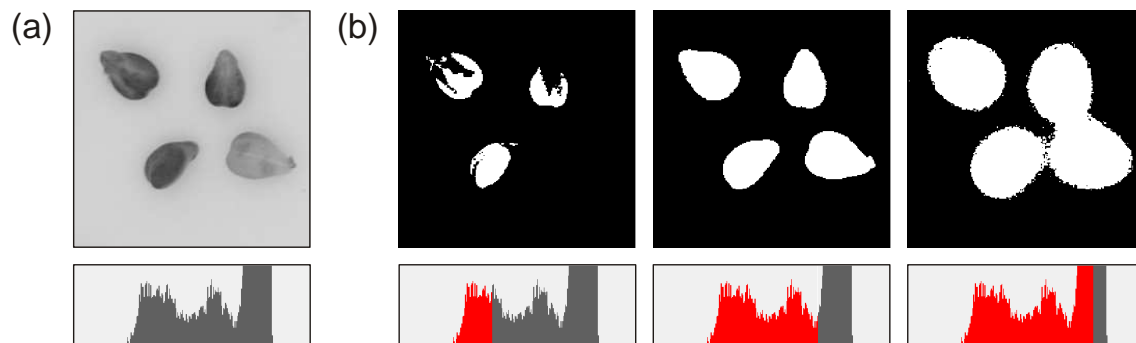


Figura 16. (a) Histograma de la imagen de semillas de uva en escala de grises. (b) Máscara de segmentación según el umbral seleccionado. El pico perteneciente al fondo aparece truncado al ampliar verticalmente el histograma.

Las imágenes de objetos similares y adquiridas bajo las mismas condiciones tendrán siempre los mismos umbrales para la segmentación. En el caso contrario, se puede construir un algoritmo de segmentación que automáticamente localice los mínimos relativos del histograma, puntos donde se encontrarán los umbrales entre la región de interés y el fondo.

Otra técnica es la segmentación *basada en los bordes*. En este caso, se realiza un barrido en toda la imagen y se buscan puntos de alta discontinuidad. La unión

de todos los puntos donde existen esas discontinuidades formará la silueta de la región de interés (línea amarilla de la Figura 17).

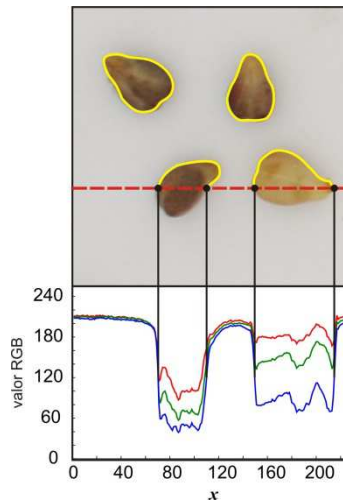


Figura 17. Segmentación basada en bordes. En la parte inferior se muestra el perfil RGB de la línea discontinua roja de la imagen.

### 3.3.2. Transformaciones morfológicas

Estas transformaciones cambian la forma y estructura de los objetos en la máscara de segmentación. Por ejemplo, una *dilatación* aumenta la máscara de segmentación al actuar en todos los píxeles frontera entre la región de interés y el fondo. Se llama *erosión* al proceso que rebaja la máscara de segmentación del mismo modo que lo hace la dilatación (Figura 18). Estas transformaciones se aplican mediante máscaras que pueden tener diferentes tamaños y formas según el efecto que se quiera conseguir, aunque normalmente tienen forma redonda.

Dilatación y erosión no son operaciones inversas, ya que al aplicarlas secuencialmente no se obtiene la forma inicial. La aplicación de una erosión y una dilatación posterior se llama "apertura", y se denomina "cierre" si se invierte el orden. El cierre y la apertura se suelen emplear para suavizar bordes, extraer formas determinadas en un entorno con ruido y eliminar salientes estrechos entre otros propósitos.

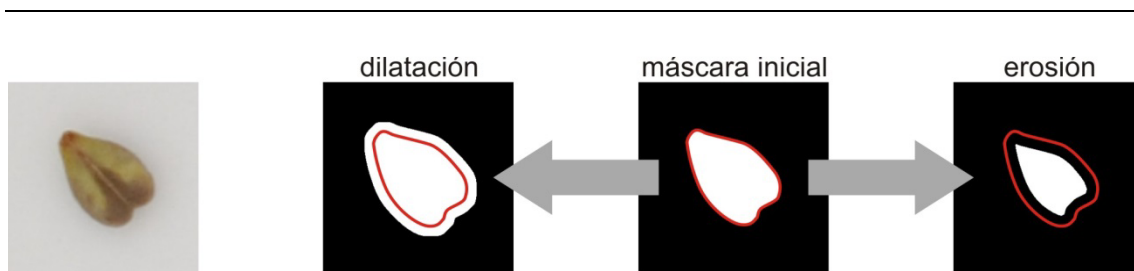


Figura 18. Transformaciones de dilatación y erosión a partir de una máscara de segmentación. La silueta en rojo señala la región de interés.

La operación de erosión puede ser útil en la medida del color. La limitación en la resolución de las imágenes hace que los objetos no estén completamente definidos. Esto hace que el fondo afecte en el color de los píxeles en el borde del objeto. Por ello es aconsejable aplicar una pequeña erosión a la máscara de segmentación para que todos los píxeles que computen en la medida del color pertenezcan claramente a la región de interés. Esto no afecta a las características morfológicas del objeto, pues se pueden utilizar máscaras diferentes para el análisis morfológico y colorimétrico.

### 3.4. Análisis morfológico de imágenes digitales

A partir de la forma del objeto se pueden medir muchas características morfológicas que pueden ser útiles en el análisis de imagen. Todas estas características se pueden extraer directamente de la máscara de segmentación, y no es necesaria la imagen original. Merece la pena destacar las medidas más comunes:

- Área: Es un número escalar que se obtiene por el simple conteo de los píxeles que pertenecen al objeto. Mediante una calibración espacial se puede obtener la correspondencia entre píxeles y unidades de longitud al cuadrado.
- Perímetro: Longitud del contorno del objeto. Como la medida se realiza con la matriz de puntos que forma la imagen, dos píxeles contiguos horizontal o verticalmente contribuirán a ese perímetro con la longitud del píxel. Si los píxeles están conectados por sus esquinas, la longitud del píxel se

multiplicará por  $\sqrt{2}$ , que es la relación que guarda el lado de un cuadrado con su diagonal.

- Redondez: Indica la similitud que tiene objeto con un círculo perfecto y se calcula mediante la siguiente ecuación:

$$\text{Redondez} = \frac{\text{perímetro}^2}{4 \times \pi \times \text{área}} \quad \text{Ec. 25}$$

Para un círculo perfecto tendrá valor 1, y se hará mayor conforme se aleje de esta forma. Algunos autores utilizan también el inverso del cociente anterior.

- Diámetro equivalente: Es el diámetro que tendría un círculo de la misma área que el objeto.
- Diámetro de Feret: Distancia que existe entre dos tangentes al objeto, paralelas entre sí y perpendiculares al eje de medida. Hay que definirlo indicando el ángulo de medida. Las Figura 19a y b muestran el diámetro de Feret medido a  $0^\circ$ ,  $90^\circ$  y a un ángulo cualquiera  $\beta$ . A partir de estas medidas se puede calcular el diámetro máximo, mínimo y medio de un objeto. Normalmente se considera el largo de un objeto su diámetro mayor y el ancho el diámetro medido perpendicularmente al diámetro mayor.
- Elipse equivalente: Es una elipse con la misma área y proporción que la imagen original. Suaviza las posibles irregularidades en los bordes de los objetos y en ocasiones es útil medir esta elipse en lugar de la figura original. Figura 19c. El cociente entre los ejes mayor y menor de esa elipse define la relación de aspecto, e indica cuan alargado es el objeto.

Aparte de los parámetros morfológicos que dan información sobre la forma del objeto, existen otros relacionados con la posición dentro de la imagen, como son el centroide del objeto y las coordenadas de la caja que contiene el objeto Figura 19d. Esta localización de objetos reduce la carga computacional de los algoritmos desarrollados en análisis de imagen. Merece la pena comentar el sistema de coordenadas en las imágenes digitales, donde el origen se encuentra en la parte superior izquierda y los píxeles toman valores positivos hacia la derecha (x, columnas) y abajo (y, filas).



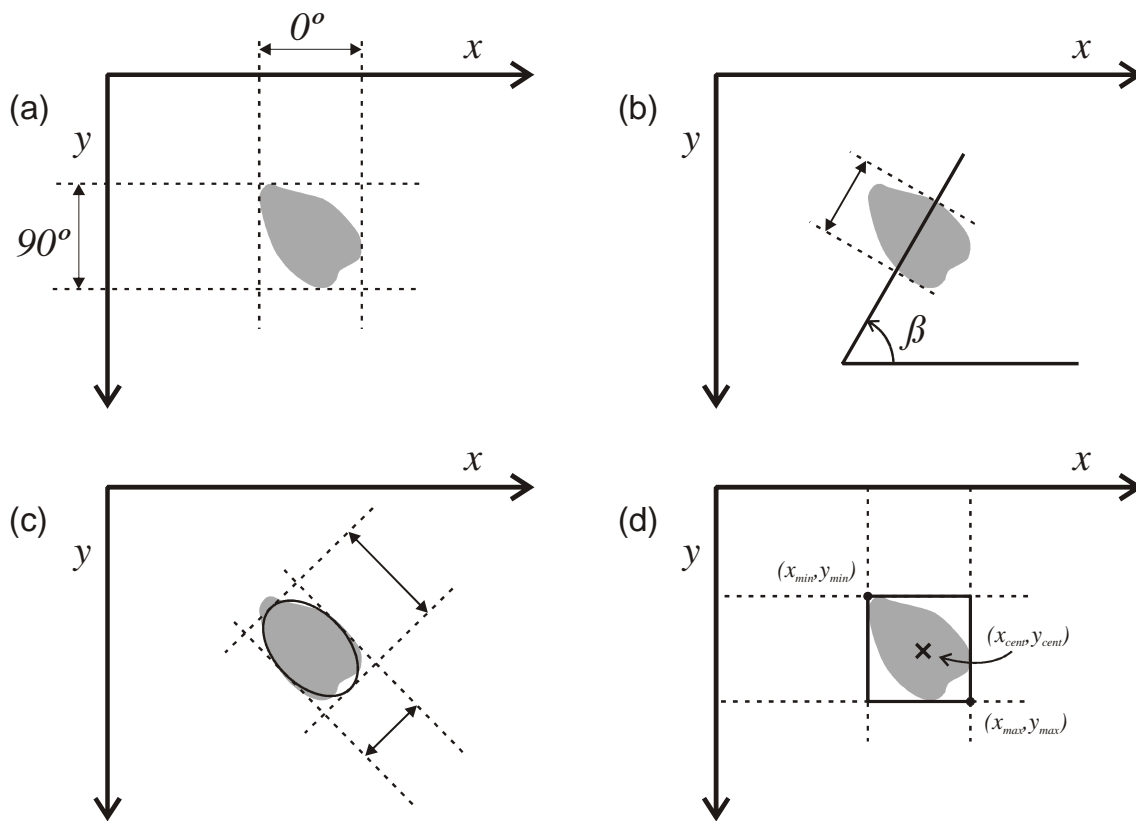


Figura 19. Medidas morfológicas sobre la máscara de una semilla de uva

### 3.5. Análisis colorimétrico de imágenes digitales

La medida del color se realiza normalmente con instrumentos que miden sobre una superficie concreta, ya sea obteniendo directamente los valores triestímulo o analizando el espectro visible y calculando los valores triestímulo a partir de él mediante las directrices de la CIE (Hutchings, Luo, and Ji, 2002). En ocasiones es necesaria la evaluación de objetos cuyo color no es homogéneo, o puede suceder que el área del objeto a evaluar sea demasiado pequeña comparada con el área de medida de la mayoría de los equipos colorimétricos.

La alta fidelidad en la calidad de imagen ha hecho que el uso de cámaras digitales para la medida del color se haya desarrollado y extendido rápidamente. Gracias a una cámara digital, la medida del color va más allá, y permite evaluar (cuantificar) aspectos relacionados con la apariencia y textura, atributos hasta

hace poco relacionados únicamente con la percepción humana (Jackman et al., 2009; Savakar & Anami, 2009; Zheng and Sun, 2008).

Sin embargo, es necesario conocer las limitaciones del uso de cámaras digitales para la evaluación del color. Cámaras digitales, televisores, pantallas LCD, programas informáticos, etc., suelen utilizar el espacio de color RGB. Si bien la CIE trata de recomendar espacios de color que promuevan su comprensión independientemente del instrumento utilizado para mejorar la reproducibilidad de las medidas, el espacio de color RGB depende del dispositivo con el que se adquiera o donde se muestre, y por ello no se debe usar como tal para medidas colorimétricas absolutas (Yam & Papadakis, 2004). Uno de los perfiles de color más utilizados es el Adobe RGB (1998), creado especialmente para reproducir el espacio de color CMYK usado en impresión. Como se muestra en la Figura 20, este espacio de color no comprende todo el espacio CIE 1931-(x,y), esto quiere decir que existen colores que no pueden ser reproducidos con cámaras o pantallas convencionales. Aparte de la dependencia del dispositivo, otra particularidad del uso de cámaras digitales como colorímetros es la falta de linealidad entre los valores RGB y los valores triestímulo XYZ. La imposibilidad de alternar de un espacio a otro con una matriz de cambio de espacio vectorial se subsana mediante diferentes ajustes no lineales. Este proceso, llamado *corrección gamma*, se puede abordar de una manera sencilla mediante la elevación a una potencia de los valores colorimétricos normalizados obtenidos con la cámara (Finlayson & Xu, 2003). A pesar de las limitaciones que supone la dependencia del dispositivo, algunos autores siguen usando las coordenadas RGB para la caracterización colorimétrica manteniendo siempre las mismas condiciones de medida (Du and Sun, 2008; Foca et al., 2011; Pacifici, 2012; Wang & Sun, 2001).

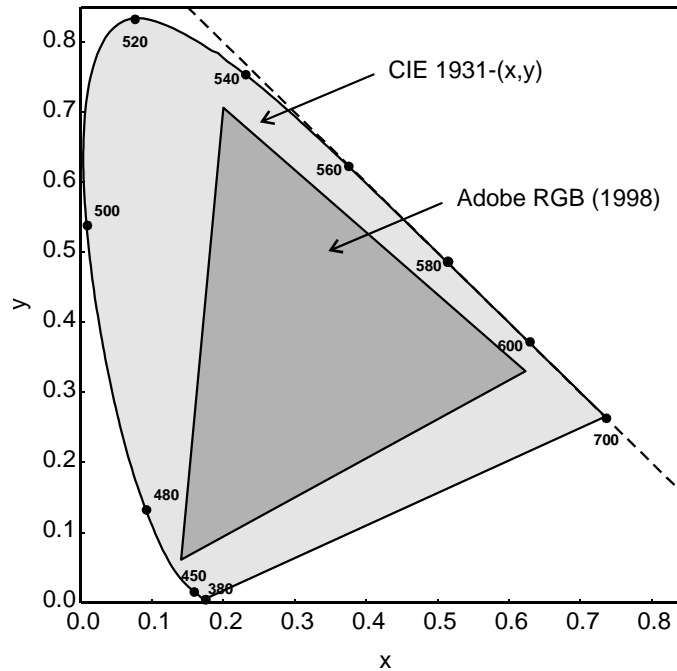


Figura 20. Espacio de color Adobe RGB (1998) comprendido como parte del diagrama CIE 1931-(x,y).

### 3.5.1. Obtención de las coordenadas CIELAB a partir del espacio RGB

Como se abordó en la sección 2 de esta introducción, el color depende de la distribución espectral de tres elementos: la fuente de iluminación, la muestra y el observador. Según esto, el color de la señal RGB de una cámara digital puede expresarse como:

$$\begin{aligned}
 R &= \int_{380nm}^{770nm} S(\lambda)r(\lambda)\bar{r}(\lambda)d\lambda \\
 G &= \int_{380nm}^{770nm} S(\lambda)r(\lambda)\bar{g}(\lambda)d\lambda \\
 B &= \int_{380nm}^{770nm} S(\lambda)r(\lambda)\bar{b}(\lambda)d\lambda
 \end{aligned}
 \tag{Ec. 26}$$

donde  $S(\lambda)$  describe la distribución espectral de una fuente de iluminación dada y  $r(\lambda)$  es la reflectancia espectral del objeto. La sensibilidad espectral de la cámara viene expresada por  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$  y  $\bar{b}(\lambda)$  para cada uno de los estímulos. Es obvio que la señal de color producida en una cámara digital depende de la sensibilidad

espectral de su sensor, y cada cámara tendrá propiedades de sensibilidad diferentes.

Para un ojo humano, los valores triestímulo X, Y y Z se calculan a partir de la intensidad espectral según las ecuaciones:

$$\begin{aligned} X &= \int_{380nm}^{770nm} S(\lambda)r(\lambda)\bar{x}(\lambda)d\lambda \\ Y &= \int_{380nm}^{770nm} S(\lambda)r(\lambda)\bar{y}(\lambda)d\lambda \\ Z &= \int_{380nm}^{770nm} S(\lambda)r(\lambda)\bar{z}(\lambda)d\lambda \end{aligned} \quad \text{Ec. 27}$$

donde  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$  y  $\bar{z}(\lambda)$  se refieren a las funciones de igualación para un observador estándar. Si dos estímulos de color tienen los mismos valores triestímulo se verán iguales en las mismas condiciones. Por ello los valores triestímulo CIE son independientes del dispositivo.

De las ecuaciones anteriores se deduce que los valores RGB de una cámara digital no coinciden con los valores CIE XYZ, ya que por lo general, los sensores de una cámara no tienen la misma sensibilidad espectral que los observadores estándares. El proceso por el cual se correlacionan las señales RGB de una cámara con los valores CIE XYZ se denomina *caracterización de la cámara*, y para ello se necesitan materiales de referencia. Existen dos tipos de caracterización, espectral y colorimétrica

Para la caracterización espectral, se mide directamente la respuesta de la cámara a una fuente monocromática en intervalos de 10 nanómetros. Este método permite medir la sensibilidad espectral de la cámara y se puede relacionar con las funciones de igualación CIE. Este tipo de caracterización se aborda con detalle en la norma internacional ISO 17321-1 (1999).

La caracterización colorimétrica se realiza mediante patrones de color. Estos patrones, cuyos valores triestímulo XYZ son conocidos, se registran con la cámara digital a caracterizar para medir sus valores RGB bajo ciertas condiciones de iluminación fijas. De los patrones comerciales disponibles, la carta de color

---

Macbeth ColorChecker (McCamy et al., 1976) es la más utilizada. Existen tres tipos de caracterización colorimétrica: mediante ajuste por mínimos cuadrados utilizando transformaciones lineales y no lineales (Kang, 1992), mediante una tabla de datos y usando interpolación (Hung, 1991; 1993) o mediante el uso de redes neuronales (Kang & Anderson, 1992). A continuación se describe un ejemplo de ajuste por mínimos cuadrados utilizando transformaciones lineales y no lineales:

Supongamos que tenemos una carta de referencia con  $N$  muestras de color. Se construye un polinomio de nueve sumandos para cada muestra como se indica a continuación:

$$\begin{aligned}
 X_1 &= a_{1,1}R_1 + a_{1,2}G_1 + a_{1,3}B_1 + a_{1,4}R_1^2 + a_{1,5}G_1^2 \\
 &\quad + a_{1,6}B_1^2 + a_{1,7}R_1G_1 + a_{1,8}G_1B_1 + a_{1,9}R_1B_1 \\
 X_2 &= a_{1,1}R_2 + a_{1,2}G_2 + a_{1,3}B_2 + a_{1,4}R_2^2 + a_{1,5}G_2^2 \\
 &\quad + a_{1,6}B_2^2 + a_{1,7}R_2G_2 + a_{1,8}G_2B_2 + a_{1,9}R_2B_2 \\
 &\quad \dots \\
 X_N &= a_{1,1}R_N + a_{1,2}G_N + a_{1,3}B_N + a_{1,4}R_N^2 + a_{1,5}G_N^2 \\
 &\quad + a_{1,6}B_N^2 + a_{1,7}R_NG_N + a_{1,8}G_NB_N + a_{1,9}R_NB_N
 \end{aligned} \tag{Ec. 28}$$

y de la misma manera para  $Y$  y  $Z$ .

En la ecuación Ec. 28,  $R_i$ ,  $G_i$  y  $B_i$  corresponden a la señal del detector de la cámara para la muestra  $i$ , y  $a_{ij}$  son los coeficientes a obtener mediante la caracterización. Si reagrupamos:

$$M = \begin{bmatrix} X_1 & X_2 & \dots & X_N \\ Y_1 & Y_2 & \dots & Y_N \\ Z_1 & Z_2 & \dots & Z_N \end{bmatrix}, \quad A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,9} \\ a_{2,1} & a_{2,2} & \dots & a_{2,9} \\ a_{3,1} & a_{3,2} & \dots & a_{3,9} \end{bmatrix}, \quad V = \begin{bmatrix} R_1 & R_2 & \dots & R_N \\ G_1 & G_2 & \dots & G_N \\ B_1 & B_2 & \dots & B_N \\ R_1^2 & R_2^2 & \dots & R_N^2 \\ G_1^2 & G_2^2 & \dots & G_N^2 \\ B_1^2 & B_2^2 & \dots & B_N^2 \\ R_1G_1 & R_2G_2 & \dots & R_NG_N \\ G_1B_1 & G_2B_2 & \dots & G_NB_N \\ R_1B_1 & R_2B_2 & \dots & R_NB_N \end{bmatrix}$$

siendo  $A$  la matriz de coeficientes,  $M$  la matriz XYZ y  $V$  la matriz RGB, se cumple que:

$$M=AV \tag{Ec. 29}$$

Como las matrices  $M$  y  $V$  se conocen, la matriz de coeficientes  $A$  puede calcularse mediante:

$$A=MV^T(VV^T)^{-1} \tag{Ec. 30}$$

Una vez que se conoce la matriz de coeficientes  $A$ , los valores XYZ pueden ser calculados a partir de los datos RGB mediante la Ec. 31:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = AV \tag{Ec. 31}$$

donde  $v^T=(R,G,B,R^2,G^2,B^2,RG,GB,RB)$ .

En principio, cuanto más alto sea el grado del polinomio usado, más precisa será la transformación entre el espacio de color. Además, La calidad de la carta de calibración, el número de colores usados y su distribución en el espacio de color afectarán a la calidad de la caracterización de la cámara. El error entre los valores observados y predichos se expresan mediante las fórmulas de diferencias de color estudiadas en la sección 2.5.

---

### 3.6. Análisis de imágenes hiperespectrales

Durante muchos años, las técnicas ópticas han experimentado un gran avance debido a su alto potencial para evaluar e inspeccionar la calidad de los alimentos de manera no destructiva, rápida y automatizable. Por un lado, los sistemas de visión artificial equipados con cámaras convencionales permiten evaluar las características de apariencia de las muestras tales como el color, forma, tamaño, textura, etc. Sin embargo, no permiten identificar o detectar fácilmente las propiedades químicas de un producto. Por otro lado, el uso de la espectroscopía convencional está ampliamente extendido para evaluar las propiedades químicas de productos de la industria alimentaria, pero su gran restricción es que la medida se realiza sobre una superficie o volumen concreto, y es necesaria cierta homogeneidad en la muestra para que el espacio de medida no sea una limitación. Las técnicas hiperespectrales surgen de la integración de la visión artificial convencional y de las técnicas espectroscópicas, determinando en un mismo análisis las propiedades espaciales y químicas de una muestra.

Nahum Gat (1998) presentó en una conferencia de la Sociedad Americana de Ingenieros Agrícolas (ASAE) los fundamentos de las imágenes hiperespectrales y el potencial que podrían tener al aplicarlas en el campo de la agricultura y en la ciencia de los alimentos en general. La técnica empezó a expandirse rápidamente y años más tarde, el *Journal of Food Engineering* publica un número especial editado por Da-Wen Sun (2004) con una revisión de todas las aplicaciones actuales de la visión artificial en la industria alimentaria. Entre los artículos publicados se puede destacar el trabajo de Mehl (2004), que utiliza las imágenes hiperespectrales para detectar defectos y contaminación en la superficie de manzanas que aparentemente serían indetectables, ya que el intervalo de longitudes de onda utilizado suele ampliarse hacia el infrarrojo cercano (NIR, del inglés *near infrared*). En la actualidad, esta técnica está ampliamente extendida y se utiliza para conocer la composición química de cereales (Nansen et al., 2008) y frutos (Sugiyama et al., 2010).

El concepto de imagen hiperespectral está alejado de la vida cotidiana, pero sus aplicaciones resultan muy beneficiosas para estudios y análisis. Una imagen en color común tiene tres bandas o canales, roja, verde y azul; y la combinación de éstas produce la formación del resto de colores. Las imágenes hiperespectrales tienen tantas bandas como se desee, y cada banda pertenece a una longitud de onda diferente, independientemente de la región del espectro electromagnético. Con ellas se logra llegar donde el ojo humano o una cámara convencional puede. Así, una imagen hiperespectral está compuesta por una batería de imágenes de una misma escena, donde cada imagen (representada normalmente en escala de grises) corresponde a la reflectancia de la muestra a una única longitud de onda. Cada píxel de la imagen contiene un espectro completo de esa posición, que queda archivado como una huella dactilar. La batería de imágenes que componen estas imágenes se conoce normalmente como cubo hiperespectral o hipercubo. En el cubo, dos de las dimensiones corresponden a las coordenadas espaciales  $x$  e  $y$ . La tercera dimensión corresponde a la longitud de onda ( $\lambda$ ) (Koehler IV et al., 2002). La dimensión espectral aporta información de qué compuestos químicos y en qué concentración se encuentran (Gowen et al., 2007). Las imágenes hiperespectrales se desarrollaron originariamente para la detección remota desde satélites, pero fue en la pasada década cuando desarrolló todo su potencial como herramienta analítica en el campo de la agricultura, astronomía, industria farmacéutica y medicina. En ciencia de los alimentos se utiliza para detectar enfermedades en frutos y cultivos, así como para conocer la composición química y el grado de maduración.

Las principales ventajas del uso de imágenes hiperespectrales para el análisis de alimentos pueden resumirse en las siguientes:

- No se necesita preparación de muestra.
- Es un método que no requiere reactivos químicos, por lo que es respetuoso con el medio ambiente, elimina la producción de residuos y reduce el riesgo de sufrir accidentes durante el análisis.
- Una vez que se construye el modelo de calibración y se valida, es un método de análisis extremadamente simple.



- 
- Es un método no invasivo y no destructivo, por lo que se puede utilizar la misma muestra para otros propósitos y análisis.
  - Aunque el equipamiento tiene un coste considerable, es económico a largo plazo comparado con los métodos tradicionales debido al ahorro en tiempo de análisis, uso de reactivos y gestión de residuos.
  - En lugar de medir un espectro ponderado como en espectroscopía convencional, se registra un espectro completo de cada punto de la imagen.
  - Posee la flexibilidad de elegir cualquier región de interés después de la adquisición de la imagen.
  - Su alta resolución espectral permite medidas cualitativas y cuantitativas.
  - Es capaz de determinar varios constituyentes simultáneamente en una misma muestra.
  - Mediante técnicas quimiométricas, se pueden producir imágenes en las cuales se represente la presencia de compuestos químicos en alguna escala de color, esta técnica se denomina "*chemical imaging*".

A pesar de las ventajas mostradas anteriormente, el uso de imágenes hiperespectrales también conlleva algunas desventajas, que se resumen a continuación:

- Las imágenes hiperespectrales están compuestas por una cantidad muy grande de información, en muchos casos redundante, y aprovechar esta información plantea muchos retos computacionales.
- Desde un punto de vista analítico, uno de los inconvenientes más grandes que presenta el uso de imágenes hiperespectrales es que es un método de análisis indirecto. Esto significa que se necesita una calibración y una validación que en ocasiones puede resultar tediosa al necesitar muchas muestras para el entrenamiento del modelo.
- Al igual que sucede con otras técnicas espectroscópicas, muchos espectros contienen bandas que se superponen y en ocasiones es difícil asignar estas bandas a grupos funcionales específicos.

- Otro factor que limita sus aplicaciones industriales radica en el requerimiento de ordenadores que puedan hacer frente a la gran cantidad de información que procesar.
- Las imágenes hiperespectrales sufren el conocido problema de la “multicolinealidad”, que obliga al uso de técnicas multivariantes como la regresión basada en el análisis de componentes principales (PCR, del inglés *principal components regression*) y la regresión basada en mínimos cuadrados parciales (PLSR, del inglés *partial least square regression*). Los efectos de la multicolinealidad pueden reducirse pero nunca eliminarse al completo por PCR y PLSR. En este aspecto, es ventajosa la selección de variables en el sentido de que no sólo se puede mejorar la capacidad predictiva del modelo de calibración, sino que también se puede simplificar el modelo eliminando la información redundante y las variables irrelevantes.

### 3.6.1. Fundamentos de espectroscopía

Las técnicas espectroscópicas comenzaron con el experimento de Isaac Newton al descomponer la luz blanca haciéndola pasar por un prisma. En la actualidad, las técnicas espectroscópicas son una eficaz herramienta para determinar los constituyentes químicos de los alimentos. Estos instrumentos utilizan elementos dispersivos que permiten evaluar la interacción entre la radiación electromagnética y la materia a diferentes longitudes de onda. El uso de una técnica espectroscópica en el infrarrojo cercano para la evaluación de la composición química requiere la acumulación de una gran cantidad de espectros que permitan realizar un proceso de calibrado. La diferencia entre el éxito y el fracaso en esta tarea radica en la calidad de los valores de referencia asociados a las muestras del conjunto de calibración. Pero, una vez se concluye el proceso de aprendizaje del método, el resultado final es un método analítico ideal.

Básicamente, los métodos espectroscópicos proporcionan *huellas dactilares* de la muestra biológica a analizar aprovechando los fenómenos de interacción de

---

la radiación con la materia, como son la reflectancia, transmitancia, fosforescencia, fluorescencia y la desintegración radiactiva. Entre las técnicas utilizadas en espectroscopía, la más utilizada en la ciencia de alimentos es la espectroscopía NIR. Las bandas que se manifiestan en este intervalo del espectro corresponden a sobretonos y bandas de combinación de las vibraciones de enlaces O-H, N-H, C-H y S-H. Por ello, la espectroscopía NIR puede aplicarse a todos los compuestos orgánicos ricos en enlaces O-H (como determinación de humedad, carbohidratos y grasas), enlaces C-H (como compuestos orgánicos y derivados del petróleo) y enlaces N-H (como proteínas y aminoácidos). En este intervalo del espectro algunas frecuencias serán absorbidas por la muestra, otras (que no corresponden con ninguna diferencia de energía entre estados vibracionales de los analitos) no se absorberán. Además, la intensidad de este fenómeno dependerá de la presencia relativa de dichos analitos. Esta compleja relación entre la intensidad de absorción y la longitud de onda constituye el espectro de absorción de una sustancia o una muestra (Pasquini, 2003).

Las técnicas de espectroscopía NIR requieren detectores de bajo ruido, un control computerizado de la adquisición de datos y el uso de algoritmos matemáticos y estadísticos multivariantes para analizar los datos. No obstante, las técnicas espectroscópicas NIR dependen sólo de la cantidad de luz reflejada o transmitida de una específica área de la muestra (según dónde se coloque el sensor), y no dan información de la distribución espacial en la muestra. Además, cuando las muestras se colocan en los espectrómetros, la homogeneidad es un punto a tener en cuenta, ya que un espectrómetro tradicional integra la información espacial, por ejemplo en el camino óptico de una cubeta. Este hecho no influye en medidas cuando la muestra es un líquido o un gas, pero en el caso de una muestra sólida (como los productos agroalimentarios), significa perder una gran cantidad de información, pues hay muchos casos en los que la localización de una característica espectral es de suma importancia.

### 3.6.2. Interacción de la luz con la muestra

La gran ventaja de la técnica está en que es una herramienta no destructiva para el análisis de alimentos basado en propiedades ópticas. Todos los materiales, incluyendo los alimentos, absorben y emiten continuamente energía que hace que se incrementen o disminuyan sus niveles energéticos. El tipo de interacción depende de la naturaleza de la materia. Fundamentalmente, cuando una onda electromagnética incide en la superficie de una muestra, la onda puede ser reflejada parcial o totalmente, y cualquier parte no reflejada penetrará en el material. Si la onda pasa a través de él sin atenuación alguna, el material se llama *transparente*, un material que atenúe parcialmente la radiación se denomina *translúcido* y un material el cual no deja pasar ninguna radiación se denomina *opaco*. La mayoría de los gases se comportan como transparentes a la radiación mientras que la mayoría de los sólidos (como son las muestras de alimentos) tienden a ser opacos. La luz visible que una muestra refleja o transmite (color) lleva información que los consumidores usan para juzgar la calidad de un producto. La limitación de la visión humana está en que solamente puede captar información de la luz visible, y existen muchas características relacionadas con la calidad de alimentos que está fuera de esta región del espectro.

En sólidos, la mayor parte de la energía penetra sólo una pequeña distancia y abandona la muestra cerca del punto de entrada. Pero parte de la radiación penetra más adentro en los tejidos alterando los estados energéticos de las moléculas que forman la muestra antes de abandonarla. Este tipo de reflectancia se llama *reflectancia difusa* y contiene mucha información química de la muestra (Abbott, 1999). La cantidad de materia que es capaz de penetrar la radiación depende de la naturaleza de la muestra y de la longitud de onda. Normalmente decrece conforme la radiación es menos energética (mayor longitud de onda).

En este aspecto merece la pena destacar la importancia de la iluminación de la muestra. Debe emplearse una fuente estable y continua en el intervalo del espectro utilizado. En imágenes hiperespectrales NIR se suelen utilizar lámparas halógenas, que proporcionan una iluminación estable en todo el infrarrojo. El

---

número de lámparas y su disposición dependerá de la geometría del equipo y de las características morfológicas de la muestra.

### 3.6.3. Terminología

A continuación se describirán algunos términos y expresiones básicos en imágenes hiperespectrales. Algunos de ellos son comunes a los términos utilizados normalmente en espectroscopía.

- *Intervalo espectral*: Describe las regiones del espectro que abarca un equipo concreto en términos de longitud de onda. Los equipos comunes suelen abarcar el visible (380–800 nm) o el visible y parte del infrarrojo cercano (400–1000 nm). Los equipos más modernos han desplazado su intervalo útil completamente al infrarrojo cercano (900–1700 nm), donde existe más información química y ofrece más posibilidades. La limitación en el material de los detectores impide que existan equipos que abarquen amplios intervalos de espectro.
- *Resolución espectral*: Está relacionado con el poder dispersivo de la radiación. La resolución espectral se define como la capacidad de discernir entre dos longitudes de onda adyacentes y se expresa en nanómetros.
- *Resolución espacial*: Determina el tamaño del objeto más pequeño que puede ser observado en la superficie de una muestra por el sensor del equipo. En otras palabras, se define como el área de la escena que es representada por un píxel de la imagen. No depende únicamente del equipo, pues también influye la distancia entre la óptica de la cámara y la muestra.
- *Número de bandas*: Es uno de los principales parámetros que caracterizan los equipos de imágenes hiperespectrales. Según sea el equipo multiespectral o hiperespectral, el número puede variar de unos pocos (normalmente menos de 10) en sistemas multiespectrales a 100–250 en equipos hiperespectrales. En equipos multiespectrales, las bandas no tienen que ser correlativas, sino que son las más útiles para una aplicación

específica. En los equipos hipertespectrales las bandas que los componen sí suelen ser correlativas.

- *Ancho de banda*: Es un parámetro que se define como la anchura a mitad de pico de una línea espectral, e indica la banda espectral más fina que se puede detectar con ese equipo.
- *Relación señal/ruido* (SNR, del inglés *signal-to-noise ratio*): Es el cociente entre la radiancia medida por el detector perteneciente a un analito y la perteneciente al ruido de los componentes electrónicos. En otras palabras, la relación señal-ruido compara el nivel de una señal deseable con el nivel del ruido de fondo. En imágenes hiperespectrales, depende de la longitud de onda, y suele haber un decrecimiento conforma la longitud de onda es mayor.
- *Huella espectral*: Las técnicas espectroscópicas aprovechan el hecho de que debido a las diferentes propiedades químicas y estructura física, los materiales refleja, dispersan, absorben o emiten radiación electromagnética siguiendo diferentes patrones a longitudes de onda específicas. Esta característica se llama huella espectral, o simplemente espectro. Cada elemento de la imagen (píxel) en una imagen hiperespectral tiene su propio espectro. La huella espectral se define como el patrón de reflectancia, absorbancia o transmitancia que posee un material a lo largo del intervalo espectral considerado, y permiten caracterizar su composición química y su estructura física (Shaw & Manolakis, 2002).
- *Imagen hiperespectral*: Está formada por muchas imágenes congruentes que representan las intensidades a diferentes longitudes de onda de una imagen bidimensional, formando así un hipercubo. Cada píxel de ese hipercubo se denomina *voxel* (en inglés, abreviatura de *vector pixel*). Al igual que sucede en imágenes digitales convencionales, cada vóxel puede adoptar valores que depende de la profundidad de bits del hipercubo. Los equipos comerciales suelen tener 8, 12, 14 o 16 bits de profundidad, aunque un equipo de 12 bits suele ser adecuado para la mayoría de las aplicaciones en análisis de alimentos. La Figura 21 muestra un ejemplo de cómo está formada una imagen hiperespectral y se puede ver la relación entre las dos dimensiones

espaciales y la dimensión espectral. De todos los píxeles, se han representado los espectros de un fragmento de semilla (rojo) y de un fragmento del fondo (azul). Además se ha extraído y marcado en amarillo la imagen que pertenece a 1216 nm. A esta longitud de onda, las semillas tienen más reflectancia que el fondo, por eso en la imagen a la derecha se ven más claras y en la gráfica de los espectros la línea azul está por encima de la roja en la línea marcada.

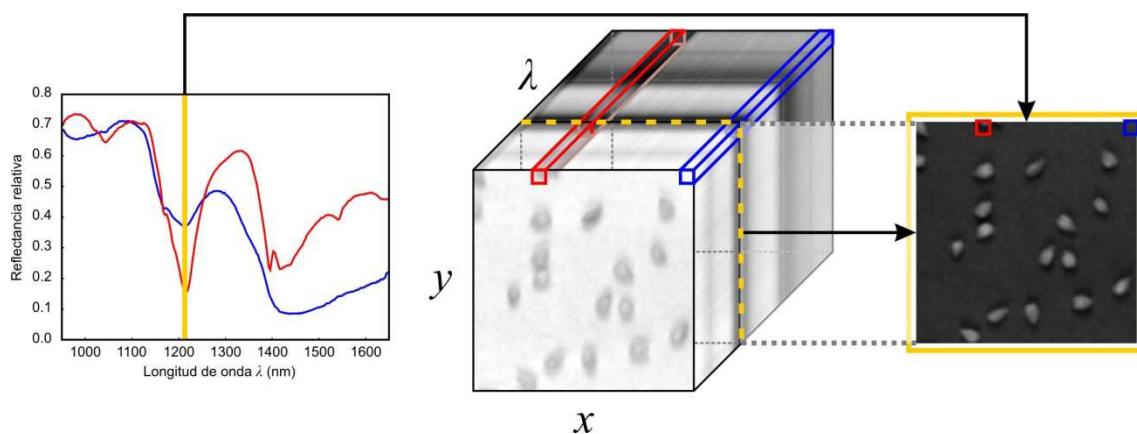


Figura 21. Diagrama esquemático de una imagen hiperespectral (hipercubo) de un fragmento de muestra de semillas de uva.

### 3.6.4. Tratamientos espectrales

El detector de una cámara hiperespectral almacena la intensidad de cada vóxel (del inglés *volumetric pixel*) en números enteros dependiendo de la profundidad de bits del equipo y que representan la reflectancia de la muestra en un punto del espacio y a una determinada longitud de onda. Esta medida no tiene interés analítico, y es por ello por lo que se utilizan normalmente unidades de reflectancia relativa mediante la *calibración de la reflectancia* de las imágenes. Para este proceso se necesitan dos medidas más aparte de la medida de la muestra. La primera de ella corresponde con un blanco que corresponde al 100% de reflectancia. Se utiliza para ello un material estándar de referencia. La otra medida se realiza con las lámparas apagadas y con la tapa de la cámara puesta. Esta medida corresponde al 0% de reflectancia y elimina el ruido de fondo de las

medidas de la muestra y del blanco. La medida de reflectancia relativa se realiza mediante la siguiente fórmula:

$$R = \frac{R_0 - B}{W - B} \quad \text{Ec. 32}$$

donde  $R_0$  es la reflectancia sin corregir,  $B$  es la reflectancia de la corriente oscura y  $W$  es la reflectancia del blanco. La reflectancia en la imagen calibrada tendrá valores comprendidos entre 0 y 1.

Para la predicción de características químicas a partir de los datos espectrales se suelen convertir los datos de reflectancia a absorbancia relativa mediante la siguiente ecuación:

$$A = \log_{10} \frac{1}{R} = -\log_{10} R \quad \text{Ec. 33}$$

Aún después de convertir los datos espectrales, las imágenes suelen presentar algunos problemas provocados por fenómenos de dispersión de la luz (*scattering*), originados por las características físicas de la muestra (como puede ser el tamaño de partícula, forma o compactación), a factores externos como la temperatura o humedad, y aspectos relacionados con la medida instrumental (Hernández-Hierro, 2008). Existen pretratamientos espectrales que pueden minimizar estos efectos con objeto de mejorar los modelos de predicción (Geladi et al., 1985; Williams, 1987). Los tratamientos más utilizados son:

- *Promediado de espectros*: Con esta estrategia se pretende reducir el ruido aleatorio, y por tanto aumentar la relación señal/ruido.
- *Corrección del efecto multiplicativo de la dispersión* (MSC, del inglés *Multiplicative Scatter Correction*): los efectos de la dispersión de la luz se asocian principalmente a factores físicos como puede ser el tamaño de partícula. Estos efectos pueden ser aditivos o multiplicativos, afectando a la relación del espectro con las propiedades químicas de la muestra. Mediante este tratamiento se pretende evitar que los efectos de la dispersión se impongan a las propiedades químicas de la muestra. La corrección se



---

efectúa utilizando un espectro ideal, que suele ser el espectro medio del conjunto de muestras.

- *Tipificación de la absorbancia (SNV, del inglés Standard Normal Variate)*: Al igual que el tratamiento anterior, intenta minimizar la dispersión debida principalmente a factores físicos, como puede ser el tamaño de partícula. Este tratamiento se aplica individualmente a cada espectro y por tanto no necesita el cálculo de un espectro medio o de referencia, ni rehacer los cálculos si se elimina alguna muestra del conjunto inicial.
- *Corrección de la tendencia (DT - Detrend)*: Este tratamiento se aplica de manera individual a cada espectro e intenta minimizar la curvatura de la línea base que surge como consecuencia del diferente tamaño de partícula y del distinto grado de compactación en su empaquetamiento.
- *Derivadas*: Son una herramienta muy utilizada en el procesado de los espectros NIR porque permiten disminuir sus problemas más característicos: bandas solapadas y variaciones en la línea base ocasionadas por características físicas de las muestras, como pueden ser el tamaño de partícula, empaquetamiento de muestras sólidas y otra serie de factores relacionados con las muestras y su presentación. El método para el cálculo de las derivadas emplea el método gap o segmento, utilizando la información de un segmento localizado en el espectro para calcular la derivada. Previamente se puede aplicar un suavizado sustituyendo los datos de un segmento por su valor promedio. La utilización de la primera derivada, pendiente de la curva espectral, elimina los términos constantes a todas las longitudes de onda, por lo que corrige desplazamientos verticales en el espectro. La segunda derivada corrige el cambio de pendiente de la curva espectral, y además, corrige los términos que varían de forma lineal con la longitud de onda. No se suelen utilizar derivadas superiores a la segunda ya que suponen un aumento del ruido en las medidas espectrales.

La combinación de los pretratamientos MSC, SNV y DT con posteriores suavizados y derivadas, ofrece un amplio abanico para intentar eliminar o reducir

## INTRODUCCIÓN

---

los efectos que enmascaran de alguna forma la señal apropiada para el propósito analítico.

# ***OBJETIVOS***





---

## **OBJETIVOS GENERALES**

El objetivo general de este trabajo es profundizar en el conocimiento del color y la apariencia de las semillas de uva en relación con su composición química. Se aplicarán técnicas ópticas no intrusivas, como la colorimetría triestímulo por digitalización y el análisis de imágenes hiperespectrales en el infrarrojo, para la caracterización espectral de las semillas. En concreto, se han planteado los siguientes objetivos específicos:

- Puesta a punto de la metodología para la adquisición de imágenes digitales, en semillas y en bayas de uva, y su aplicación al seguimiento de la maduración.
- Desarrollo de una metodología para la evaluación de la heterogeneidad cromática de las muestras mediante la definición de elipsoides en el espacio de color CIELAB.
- Evaluación del método de elipsoides en la estimación de la composición y el estado de maduración de las semillas de uva.
- Desarrollo de la metodología hiperespectral y aplicación a la discriminación varietal de semillas de uva.
- Estudio de la relación entre la composición química y la reflectancia espectral de las semillas de uva obtenida mediante análisis de imágenes hiperespectrales.

## OBJETIVOS GENERALES

---

---

## GENERAL OBJECTIVES

The main objective of this work is to deepen the knowledge of colour and appearance of grape seeds regarding their chemical composition. Optical techniques of digitalizing and near infrared hyperspectral image analysis will be applied for the spectral characterization of seeds. In particular, the following specific aims have been proposed:

- To set up the methodology for the acquisition of digital images of grape seeds and berries, and its application for monitoring the ripening.
- To develop a methodology for evaluating the chromatic heterogeneity of samples by defining ellipsoids in the CIELAB colour space.
- To assess the method of ellipsoids for estimating the composition and stage of ripening of grape seeds.
- To develop the hyperspectral methodology and its application for the varietal discrimination of grape seeds.
- To study the relationship between the chemical composition of grape seeds and their spectral reflectance obtained by hyperspectral imaging.







***MATERIAL Y MÉTODOS***



---

## 1. MUESTRAS

Los estudios llevados a cabo en esta Memoria se han centrado principalmente en uvas (*Vitis vinifera L.*) de viñedos localizados en la Denominación de Origen Condado de Huelva en las campañas 2009, 2010, 2011 y 2012 (Figura 22).



Figura 22. Localización de la Denominación de Origen Condado de Huelva dentro de las zonas vitivinícolas de España.

Se muestrearon cuatro fincas según disponibilidad, variedad de uva y tipo de suelo (Figura 23).

- *Finca 1 (37°16'53" N, 6°34'32" O)*: Localizada en el término municipal de Rociana del Condado, con viñas de la variedad Zalema, suelo tipo arena y cepas con poda en vaso. La variedad de uva blanca Zalema es autóctona del Condado de Huelva y de la comarca del Aljarafe sevillano. *Zalemo* y *Rebazo* son sinonimias de Zalema. La planta de Zalema tiene un porte de vegetación erguido, su hoja joven es amarilla cobriza y algodonosa en el envés. Produce racimos grandes, con pedúnculos cortos y poco lignificados. Las bayas son de tamaño medio, circular, con piel basta de color verde-amarillento y pulpa carnosa con dos semillas por término medio. Su ciclo vegetativo es de brotación a principios de marzo, la floración a mediados de mayo y el envero sobre el cambio de mes de julio a agosto. Su maduración suele completarse durante la primera semana de septiembre y pierde las hojas en el mes de noviembre (García de Luján et al., 1990). Por otro lado, el suelo arena contiene alrededor de un 75% de arenas finas o groseras con tendencia a formar corazas que impiden la aireación y necesitan frecuentes labores superficiales. Su permeabilidad es muy buena y dan rendimientos cuantitativos muy altos.
- *Finca 2 (37°19'42" N, 6°32'29" O)*: Localizada en el término municipal de Bollullos Par del Condado, con viñas de la variedad Zalema y poda en espaldera. Además del tipo de poda, difiere de la finca 1 en el tipo de suelo. Esta finca tiene suelo de tipo barro (arcillo-arenosos). Corresponde a suelos rojos mediterráneos (46.8% de arcilla y 42.6% de arenas finas), de textura pesada y escasa permeabilidad. Son menos favorables a la vid que los suelos arenosos.
- *Finca 3 (37°19'8" N, 6°35'55" O)*: Localizada en el término municipal de Rociana del Condado, con suelo tipo barro y poda en espaldera. La variedad muestreada en esta finca fue la Tempranillo. Esta variedad tinta es la más cultivada en España. Aunque su origen se encuentra en la zona del Ebro, su cultivo se ha extendido actualmente por el resto de regiones españolas, siendo una variedad de producción media-alta, que se adapta a cualquier suelo y es bastante resistente a enfermedades. En cuanto a su comportamiento agronómico, resulta fácil de cultivar, dando lugar a vinos

de mucha calidad, con escaso poder oxidativo, bien equilibrados y aromáticos, de acidez y graduación medias y de característico color rojo rubí (Liu et al., 2006).

- *Finca 4* ( $37^{\circ}20'41''$  N,  $6^{\circ}31'4''$  O): Localizada en el término municipal de Bollullos Par del Condado, con suelo tipo arena y poda en espaldera. La variedad de esta finca es Syrah, ampliamente implantada en centro y el sur de Francia, y debido a su buena adaptación a diferentes condiciones agroclimáticas, en las últimas décadas su cultivo se está extendiendo a otras regiones vitivinícolas por todo el mundo, especialmente en regiones de clima cálido. Los vinos elaborados a partir de ésta se caracterizan por presentar una acentuada intensidad de color, con gran cantidad de pigmentos, taninos, aromas y sabores más especiados que afrutados, así como con una gran aptitud para el envejecimiento (Robinson, 1996).

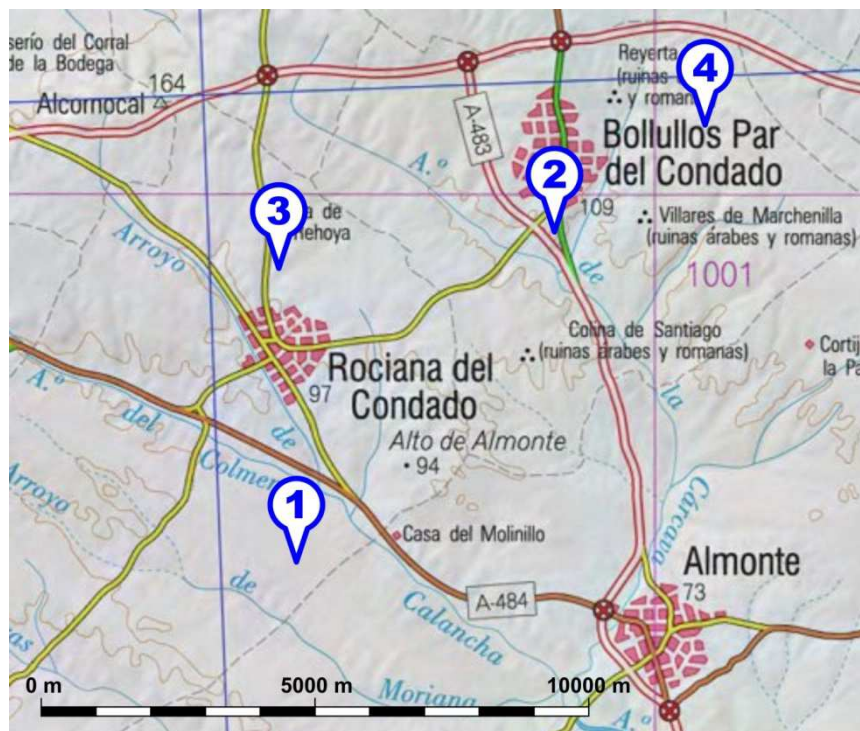


Figura 23. Situación de las fincas muestreadas.

### **1.1. Proceso de muestreo**

Se cogieron muestras desde principios de julio hasta la fecha de vendimia, que sucedió desde finales de agosto a mediados de septiembre dependiendo de la variedad y el año, y el muestreo se realizó dos veces por semana (lunes y jueves) a primera hora de la mañana. Las bayas se tomaron siempre de los mismos cuatro linios marcados en cada finca, con una separación de tres linios entre ellos y evitando el muestreo de los linios de los extremos, así como de las partes cercanas a las calles principales de la finca. De cada linio, se tomó un par de bayas de vides alternas, con la mano derecha y de todas las partes del racimo, sin mirar para evitar condicionar el muestreo en la medida de lo posible. La baya se cogió manteniendo el pedicelo para ralentizar los procesos oxidativos tras el muestreo. Al llegar al final del linio, se muestrearon las vides que quedaban enfrentadas en la misma calle, de este modo se aseguró el muestreo representativo de la parte de umbría y solana. Siguiendo este procedimiento, se recogió una muestra bruta de 2 kg y se llevó al laboratorio refrigerada hasta el momento del análisis. Esas mismas muestras fueron utilizadas para los análisis de maduración oficiales para el Consejo Regulador de la Denominación de Origen Condado de Huelva.

Para uno de los estudios se utilizaron semillas de uva de la variedad tinta Graciano pertenecientes a dos viñedos de la Denominación de Origen Calificada Rioja en la campaña de 2010.

---

## 2. ANÁLISIS DE IMAGEN

### 2.1. Adquisición de imágenes digitales

La adquisición de imágenes digitales se realizó mediante el sistema DigiEye® de VeriVide (Leicester, Reino Unido) (Luo et al., 2001). Está especialmente diseñado para medida del color según las directrices de la CIE y para la evaluación de la apariencia. El equipo está formado por una cabina de iluminación controlada, una cámara digital y un ordenador que gobierna el equipo (Figura 24).

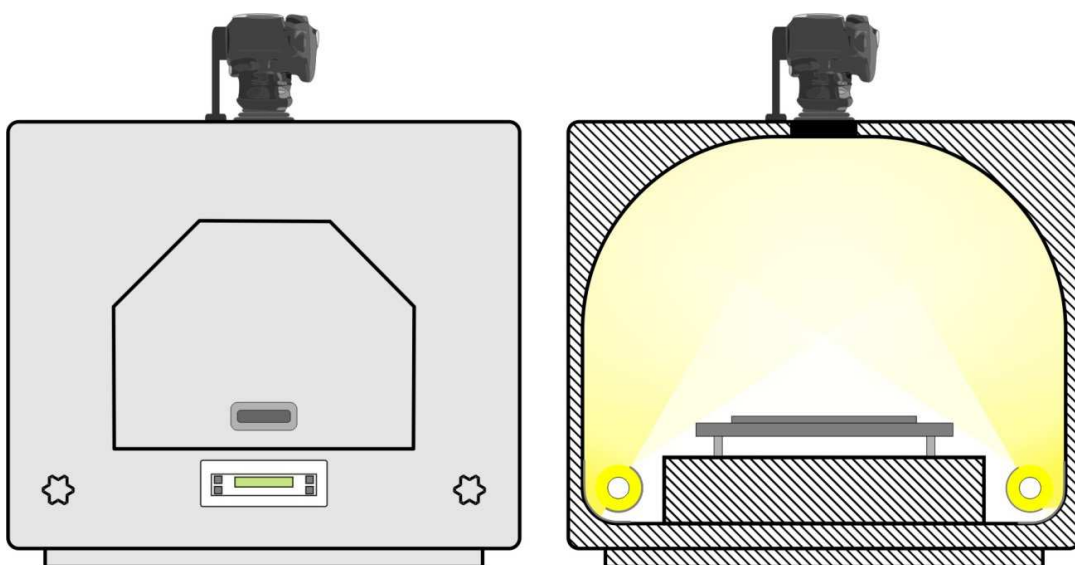


Figura 24. Esquema de la cabina del equipo DigiEye®.

### 2.1.1. Cabina de iluminación

El habitáculo de la cabina tiene una anchura de 68 cm, una profundidad de 64 cm y una altura de 40 cm. En la parte inferior, y 15 cm por debajo de la bandeja de muestra hay dos tubos fluorescentes que emulan el iluminante D65 (CIE, 2007). Dos deflectores reparten la luz de los tubos fluorescentes por la parte superior de la cabina, y es la cúpula abovedada de ésta la que ilumina la muestra de manera difusa. Este tipo de iluminación se conoce como “tipo domo”, y ofrece el máximo nivel de rendimiento en iluminación difusa, ya que asegura una iluminación homogénea desde todos los puntos posibles y se minimizan los efectos de sombras y brillos indeseables para una correcta evaluación del color.

### 2.1.2. Cámara digital

El dispositivo para la adquisición de imágenes fue una cámara tipo réflex Nikon® D80 de 10.2 megapíxeles (3872×2592 píxeles) con un objetivo Nikkor® 35 mm f/2D. Los parámetros de la cámara se fijaron de acuerdo con los valores recomendados por VeriVide. El tiempo de exposición fue 1/15 segundos, adecuado para una correcta cantidad de luz según las condiciones de la cabina, y es suficiente grande para eliminar el efecto de parpadeo de la iluminación fluorescente, la cual, lo hace con la misma frecuencia que la corriente alterna (50 Hz en Europa). Otro de los parámetros a tener en cuenta a la hora de adquirir las imágenes es la abertura del diafragma, que se expresa como un número inverso. Una abertura pequeña supondrá la entrada de poca luz en la cámara. Sin embargo, las imágenes así tomadas tendrán una gran profundidad de campo. Es decir, objetos que estén a diferentes distancias aparecerán igualmente enfocados. Por el contrario, una abertura de diafragma grande conllevará más entrada de luz en la cámara, pero la profundidad de campo será muy baja. O sea, solamente aparecerán enfocados los objetos que estén a una determinada distancia según el enfoque. En nuestras



---

imágenes, la abertura se fijó en todos los casos en  $f/6.3$ . Este valor adopta una situación de compromiso, deja pasar suficiente cantidad de luz y permite tomar imágenes de objetos que tengan cierto volumen sin que haya que preocuparse por la profundidad de campo a la hora del enfoque.

Por último, la sensibilidad de la cámara se puede ajustar mediante la ganancia electrónica del dispositivo CCD. Este parámetro se fija mediante el valor ISO, llamado así por ser el equivalente a este valor en los carretes de fotografía analógica. Ajustando un alto valor ISO se necesitará menos cantidad de luz para conseguir una señal aceptable en el dispositivo CCD. Por contra, el ruido instrumental se manifestará como una presencia de “grano” indeseable en las imágenes. Las imágenes fueron tomadas con un valor ISO 200.

## 2.2. Adquisición de imágenes hiperespectrales

La adquisición de imágenes de imágenes hiperespectrales se realizó con un equipo (Figura 25) formado los siguientes componentes:

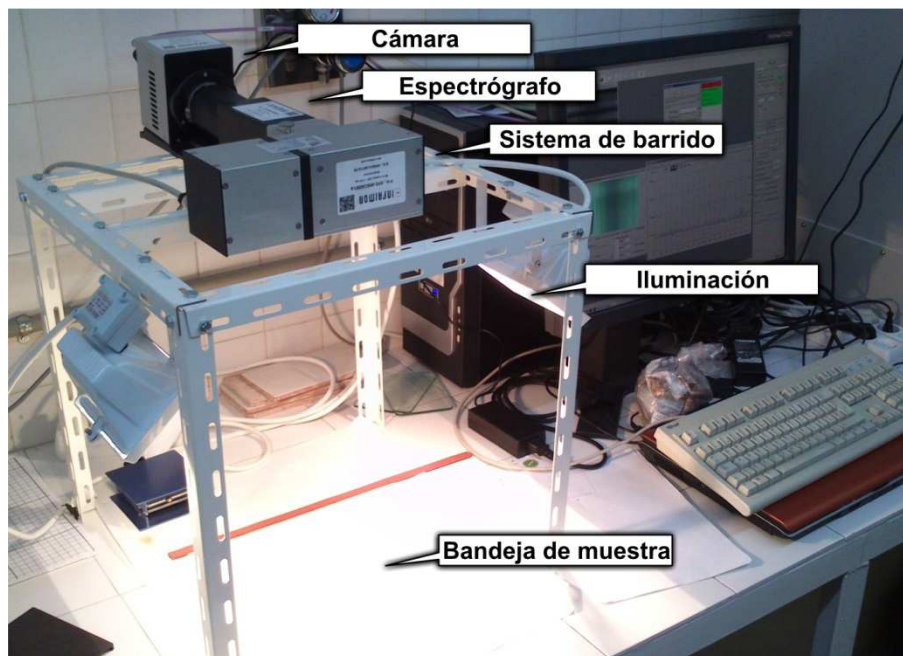


Figura 25. Cámara hiperespectral utilizada.

### 2.2.1. Cámara

Se utilizó una cámara Xeva 1.7-320, fabricada por Xenics® (Lovaina, Bélgica). Dispone de un detector tipo InGaAs con refrigeración termoeléctrica y con un intervalo espectral útil de 900 a 1700 nm, compuesto por una matriz de 320×256 píxeles. La cámara tiene una fuente de alimentación propia de 12 V y se comunica con el ordenador mediante un puerto USB 2.0.

### 2.2.2. Espectrógrafo

Como elemento dispersivo de la luz se utilizó un espectrógrafo Inspector N17E de Specim® (Oulu, Finlandia). Su óptica permite dispersar cada punto de una línea de luz (monodimensional) en longitudes de onda que van de 900 a 1700 nm, y forma una imagen bidimensional de 14.2×7.6 mm (espacial×espectral). De este modo, la imagen en su dimensión espacial queda registrada en el eje de 320 píxeles de la cámara y la dimensión espectral de 7.6 mm queda registrada en el eje de la cámara de 256 píxeles. La resolución espectral del equipo es de 3.25 nm, resultado de repartir el intervalo útil del espectrógrafo en los 256 píxeles de la cámara.

### 2.2.3. Sistema de barrido

La dimensión espacial restante se genera gracias al dispositivo de barrido Mirror Scanner, Specim® (Oulu, Finlandia). Está diseñado para ser acoplado al espectrógrafo y es útil en el mismo intervalo del espectro electromagnético. Dispone de un espejo móvil que “barre” la superficie de muestra y permite construir el hipercubo. Tiene una fuente de alimentación propia de 24 V y se comunica con el ordenador mediante un puerto USB 2.0.

---

#### 2.2.4. Iluminación

A la altura de la cámara, separadas 50 cm y dirigidas a 45° respecto a la superficie de muestra se colocaron dos lámparas halógenas tipo cuarzo iodo de 70 W Prilux® (Barberá del Vallés, Barcelona).

#### 2.2.5. Bandeja de muestra

Se utilizó como fondo una superficie de polietileno, cuyo espectro característico fue útil en los procesos de segmentación de las imágenes hiperespectrales.

#### 2.2.6. Blanco de referencia

Para normalizar los valores de referencia se utilizó una pieza cerámica certificada Spectralon® de 300×25×10 mm (Labsphere Inc., North Sutton, USA).

#### 2.2.7. Software para la adquisición de imágenes

El control de la cámara y del sistema de barrido, la sincronía de ambos dispositivos y el almacenamiento de las imágenes, se gobernó desde el software SpectralDAQ 3.62, Specim® (Oulu, Finlandia). Según las condiciones de iluminación y la potencia del ordenador, se fijó un tiempo de integración de las imágenes de 9 milisegundos y una tasa de muestreo de 50 Hz (imágenes por segundo).



---

## 3. ANÁLISIS QUÍMICO

### 3.1. Preparación de las muestras

#### 3.1.1. Material y reactivos

- Molino analítico IKA® (Alemania).
- Balanza analítica METTLER TOLEDO® AE-260 (Barcelona, España)
- Baño Ultrasonidos P-Selecta® Ultrasons-H.
- Vortex Genius 3 IKA® (Alemania).
- Centrífuga allegro X-12R Beckman Coulter®.
- Acetona (Panreac®, Ref. 361007.1612).
- Agua desionizada (Milli-Q, Millipore®).
- Ácido acético glacial (Panreac®, Ref. 141002.1211).
- Metanol calidad HPLC (Panreac®, Barcelona, España, Ref. 131091).
- Ácido tartárico (Panreac®, Barcelona, España, Ref. 141066).
- NaOH 0.5 M (J.T.Baker®, Holanda)
- Filtros de jeringa de nylon 13 mm × 0.45 µm Millipore®.
- Concentrador a vacío Eppendorf® Concentrator Plus.

### 3.1.2. Procedimiento analítico

El procedimiento para la preparación de muestra se realizó según el objeto del análisis.

Para el análisis cromatográfico, las semillas se liofilizaron y se molturaron hasta conseguir un polvo homogéneo, y se empleó en todas las muestras el mismo tiempo de molturado para minimizar la variabilidad en la extracción. Se pesaron 0.5 g de polvo de semilla con precisión de 0.001 g y se realizó una extracción con 10 mL de disolución extractante (acetona/H<sub>2</sub>O/AcOH; 70:29.7:0.3 v/v) en tubos de polipropileno de 15 mL (Falcon®, VWR International Eurolab, Barcelona), donde se agitaron durante 30 segundos. A continuación, se sometieron a ultrasonidos 5 minutos, se dejaron reposar 20 minutos y se trataron otros 5 minutos más en ultrasonidos. Tras la extracción, se separó el agente extractante del residuo, centrifugando los tubos a 3000 rpm y 25 °C durante 10 minutos. 7 mL del extracto se concentraron a vacío durante dos horas sin calefacción para evitar la polimerización y la oxidación de los fenoles, asegurando la completa evaporación de la acetona contenida en el extracto. Este punto fue muy importante, ya que durante la optimización del método y en las primeras pruebas, se detectó que una mínima cantidad de acetona provocaba que los compuestos fenólicos eluyeran al principio del cromatograma conjuntamente sin resolver los picos. Así, tras la concentración, quedó únicamente la fracción acuosa, que se llevó a un volumen de 5 mL con fase móvil A y se hizo pasar por un filtro de nylon de 0.45 µm de tamaño de poro.

La extracción para el análisis espectrofotométrico se llevó a cabo de dos maneras diferentes: extracción exhaustiva y extracción en vino sintético. Para la extracción exhaustiva, las semillas se liofilizaron y se molturaron igual que en el caso anterior. Un gramo de polvo de semilla se sometió a extracción con 10 mL de una mezcla metanol:agua (75:25), se sometieron 15 minutos a ultrasonidos y 15 minutos a centrífuga, repitiendo este proceso dos veces más. Los extractos metanólicos se agruparon y se llevaron a un volumen final con 50 mL de metanol. Para la extracción en vino sintético, dos gramos de semillas intactas se maceraron

---

en 50 mL de vino sintético durante 72 horas. El vino sintético consistió en una disolución acuosa con 12.5% de etanol, 4g·L<sup>-1</sup> de ácido tartárico y cuyo pH se ajustó a 3.6 con NaOH 0.5 M.

## 3.2. Análisis cromatográfico

La separación, identificación y cuantificación de los compuestos fenólicos se utilizaron modificaciones de los métodos descritos por Sandhu & Gu (2010).

### 3.2.1. Material y reactivos

— Cromatógrafo de líquidos de alta eficacia HPLC Agilent Technologies® HP 1200, equipado con:

- Bomba cuaternaria (G1311A).
- Desgasificador de vacío (G1322A).
- Compartimento termostatzado para columna (G1330B).
- Inyector automático (G1329A).
- Detector de diodos (G1316B).
- Columna Zorbax ODS de 4.6 250 mm y 5 um de tamaño de partícula (Agilento Technologies®).

— Agua desionizada (Milli-Q, Millipore®).

— Ácido fórmico 98% calidad HPLC (Merck®, Darmstadt, Alemania).

— Metanol calidad HPLC (Panreac®, Barcelona, España, Ref. 131091).

— Catequina (Sigma Aldrich®, Ref. C1251).

— Epicatequina (Sigma Aldrich®, Ref. 39263).

— Ácido gálico (Analyticals Carlo Elba®, Ref. 406335).

— Ácido protocatéquico (Sigma Aldrich®, Ref. 39421).

— Metanol calidad HPLC (Panreac®, Barcelona, España, Ref. 131091).

### 3.2.2. Procedimiento analítico

El método cromatográfico se configuró para que registrara la señal a 280 y 320 nm, termostatar la columna a 25 °C y fijara un flujo de fase móvil constante de 1 mL·min<sup>-1</sup>. Los disolventes usados como fase móvil fueron una disolución acuosa al 0.5% (v/v) de ácido fórmico (Fase A) y metanol puro (Fase B). Se aplicó gradiente lineal con la siguiente programación: 5% B de 0 a 2 min, de 5 a 20% B de 2 a 10 min, de 20 a 30% B de 10 a 15 min, de 30 a 35% B de 15 a 20 min, de 35 a 80% B de 20 a 60 min, de 80 a 85% B de 60 a 65 min, y de 85 a 5% B de 65 a 75 min para reequilibrar la columna para la siguiente muestra. Para realizar las rectas de calibrado, se realizó una disolución madre multipatrón y se prepararon a partir de ésta viales con distintas concentraciones. Mediante medida de los tiempos de retención y la comparación de espectros con la biblioteca se identificaron las áreas de los distintos compuestos fenólicos.

### 3.3. Análisis espectrofotométrico

Para el análisis de flavanoles se siguió una modificación del método de Vivas et al. (Vivas et al., 1994).

#### 3.3.1. Material y reactivos

- Espectrofotómetro Agilent 8453 (Agilent Technologies®, Palo Alto, EE.UU.)
- Metanol calidad HPLC (Panreac®, Barcelona, España, Ref. 131091).
- 4-(dimetilamino)cinamaldehído (DMACA) (Sigma Aldrich®, Ref. 39421)
- Ácido clorhídrico 32 % (Panreac®, Barcelona, España, Ref. 132176)
- Catequina (Sigma Aldrich®, Ref. C1251).
- Agua desionizada (Milli-Q, Millipore®).
- Micropipetas Eppendorf® de intervalos de volumen 100-1000 µL y 20-200 µL)
- Filtros de jeringa de nylon 13 mm × 0.45 µm Millipore®.



---

— Concentrador a vacío Eppendorf® Concentrator Plus.

### 3.3.2. Procedimiento analítico

10 o 20  $\mu\text{L}$  de extracto, según fuera de extracción exhaustiva o de maceración en vino sintético, se mezclaron con 190 o 180  $\mu\text{L}$  de metanol respectivamente. Se añadió 1 mL de reactivo DMACA y tras 10 minutos se midió la absorbancia a 640 nm. El reactivo DMACA se preparó inmediatamente antes de su uso, y contenía 0.1% (p/v) de 4-(dimetilamino)cinamaldehído en una disolución HCl:metanol (1:10, v/v). Los análisis se realizaron por triplicado. Los flavanoles totales se cuantificaron mediante una recta de calibrado de (+)-catequina.



---

## 4. TRATAMIENTO DE RESULTADOS

### 4.1. Procesado de imágenes

Todas las imágenes digitales en color e hipercubos fueron procesados mediante algoritmos programados bajo MATLAB R2012B (The MathWorks Inc., Natick, USA). MATLAB (en inglés, abreviatura de *Matrix Laboratory*) es una herramienta de software matemático que funciona con un lenguaje de programación propio. Su característica principal es la facilidad con la que el programa manipula las matrices. Esto resulta muy útil en análisis de imagen, ya que tanto las imágenes digitales como los hipercubos en imágenes hiperespectrales pueden ser considerados matrices tridimensionales. Dentro de MATLAB, existen paquetes de herramientas (en inglés *toolboxes*), que no son más que conjuntos de programas diseñados para que el usuario no tenga que realizar desde cero algunas operaciones básicas en algunos campos de aplicación. Dentro de este trabajo se han utilizado las *toolboxes Image Processing, Fuzzy Logic, Partial Differential Equation* y *Statistics*. Los códigos de los programas realizados para esta tesis se encuentran en el Anexo.

## 4.2. Análisis estadísticos y quimiométricos

La estadística descriptiva tiene como objetivo la descripción del colectivo estudiado, que puede ser toda la población (censo) o una parte de la misma (muestra). Cuando en una población dada se selecciona adecuadamente un grupo de individuos (muestra), la descripción de éstos es el objetivo de la estadística descriptiva (Martín, 2001). El análisis descriptivo de los datos da una visión general, complementaria a la posterior confirmación mediante el uso de técnicas estadísticas multivariantes.

Los elementos utilizados para el análisis descriptivo de los datos en este estudio son la media aritmética ( $\bar{x}$ ) y desviación estándar ( $SD$ ), y representaciones gráficas tales como diagramas de barras, gráficos de líneas para comparar dos variables o gráficos de dispersión (*scatterplots*) para estudiar la relación entre dos variables.

Por otro lado, la *International Chemometrics Society* (ICS) la define como la Quimiometría como *la disciplina química que utiliza métodos matemáticos y estadísticos para diseñar o seleccionar procedimientos de medida y experimentos óptimos, y para proporcionar la máxima información química mediante el análisis de datos químicos*. Utiliza las herramientas matemáticas, estadísticas y de lógica formal para:

- Diseñar o seleccionar el procedimiento experimental óptimo.
- Proporcionar el máximo de información analítica relevante analizando los datos químicos.
- Obtener conocimiento sobre sistemas químicos.

Los tratamientos estadísticos y quimiométricos se llevaron a cabo con los programas MATLAB R2012B (The MathWorks Inc., Natick, USA), Statistica 8.0 (StatSoft Inc., Tulsa, USA) y Unscrambler 9.7 (CAMO, Trondheim, Norway).

---

#### 4.2.1. Análisis de la varianza (ANOVA)

En general, el propósito del Análisis de la Varianza (ANOVA) es encontrar diferencias significativas entre medias, mediante la comparación de las varianzas, entre los grupos formados (Norman et al., 1996).

ANOVA es el análisis estadístico paramétrico de modelos lineales en los que la variable dependiente (respuesta) es numérica y las variables independientes (factores) son cualitativas. Los factores actúan así como criterios de clasificación de los datos. Los distintos valores que puede tomar un factor se denominan niveles. De este modo se forman grupos que se desean comparar (Martín, 2001).

El valor del nivel de significación estadística  $p$  representa un índice de descenso de la fiabilidad del resultado. El valor  $p$  representa la probabilidad de error que implica aceptar como válido el resultado observado, es decir, representativo de la población. En la aplicación del ANOVA se asume que las variables incluidas en la comparación presentan una distribución normal. Cuando los datos de la variable no se ajustan a una distribución normal, existen alternativas al ANOVA, como el test de Kruskal-Wallis para comparar diferencias entre grupos de muestras independientes, y el test de Friedman cuando se quiere comparar variables medidas en la misma muestra (Martín, 2001).

#### 4.2.2. Análisis General Discriminante (GDA)

El Análisis por Funciones Discriminantes tiene como objetivo determinar hasta qué punto es posible separar dos o más grupos de muestras previamente establecidos, y qué variables son las que mejor discriminan entre ellos (Johnson, 2000; Tabachnick & Fidell, 2012).

Para el análisis discriminante es necesario una variable cualitativa (variable dependiente) y un conjunto de variables cuantitativas y/o cualitativas dicotómicas (variables independientes). Su objetivo es estimar mediante funciones lineales, funciones discriminantes, de las variables independientes, la probabilidad de que

cada uno de los individuos pertenezca a cada uno de los grupos ya definidos por las categorías de la variable dependiente. Esta clasificación se hará mediante unos criterios o propiedades comunes de los individuos dados por las variables independientes. La función discriminante representa la mejor separación estadística entre los puntos, es decir, la máxima cantidad de varianza residual.

El método más común dentro del es el procedimiento iterativo: las variables son añadidas (“hacia delante”, *forward stepwise*) o eliminadas (“hacia atrás”, *backward stepwise*) de las funciones discriminantes una a una, realizando un estudio de discriminación después de cada paso. En el caso del método *forward stepwise*, cada inclusión viene determinada por el valor del estadístico  $F$ , que indica, para cada variable, su significación en la discriminación entre los grupos, de tal manera que se incluyen las variables que poseen un valor  $F$  superior al establecido en el análisis. De esta forma, el proceso concluye cuando no se da una discriminación significativamente mejor que la anterior.

En resumen, a partir de las variables originales, se establecen funciones que discriminan entre grupos previamente establecidos. Posteriormente, estas funciones pueden ser utilizadas en la clasificación de muestras de origen desconocido asumiendo el error correspondiente, que viene expresado por los propios porcentajes de clasificación correcta, o porcentajes de acierto, de las muestras utilizadas inicialmente en la obtención de las ecuaciones.

La correlación canónica es el valor atribuido a cada una de las funciones discriminantes extraídas. Ésta mide las desviaciones de las puntuaciones discriminantes entre grupos respecto a las desviaciones totales sin distinguir grupos. Si su valor es grande, próximo a 1, la dispersión será debida a la diferencia entre grupos y, por tanto, la función discriminará mucho (Martín, 2001).

### 4.2.3. Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales (PCA, del inglés *Principal Component Analysis*) es un método de reconocimiento de pautas no prefijadas,

---

habitual en análisis multivariante, cuyo objetivo es reducir el número de variables iniciales, sin perder información y así poder observar estructuras y tendencias en las muestras (Brereton, 1992). De forma muy básica, el PCA consiste en:

- Estimar la dimensionalidad real de los datos de las muestras estudiadas; y reducir esta dimensionalidad manteniendo la información relevante del sistema.
- Calcular un nuevo sistema de coordenadas, para lo que se buscan las direcciones que expliquen la máxima variabilidad en los datos de las muestras. Estos nuevos ejes se denominan componentes principales (PC). Así el primer componente principal es la dirección que explica la máxima variabilidad de los datos de las muestras; el segundo, se escoge que sea perpendicular al primero, y que explique la máxima variabilidad de los datos de las muestras una vez restada la explicada por el primer componente principal y así sucesivamente.
- Definir estos nuevos ejes matemáticamente, mediante los *loadings*, cargas o pesos que son los cosenos de los ángulos que forman cada uno de estos nuevos ejes con los antiguos, y los scores o puntuaciones que son las coordenadas en estos nuevos ejes. De tal forma que:

$$\text{Variable dependiente} = (\text{score}_1)(\text{loading}_1) + (\text{score}_2)(\text{loading}_2) + \dots + (\text{score}_n)(\text{loading}_n) + E \sim \text{residual}$$

En este trabajo, los análisis de componentes principales se aplicaron sobre la información espectral de las muestras, y cada variable independiente correspondió a una longitud de onda. Los componentes principales se pueden interpretar geoméricamente de tal forma que el espectro medido a  $J$  longitudes de onda, constituye un conjunto de  $J$  variables que puede ser descrito como un vector. Se puede construir un espacio de  $J$  dimensiones, de tal manera, que cada dimensión corresponda a una de las longitudes de onda, y se pueda representar el mismo objeto como un punto en ese nuevo espacio. Si tenemos  $M$  muestras, se pueden representar como  $M$  puntos en el espacio de  $J$  dimensiones. El objetivo del PCA es hallar las direcciones en que están agrupados los  $M$  puntos de los objetos en el

espacio de  $J$  dimensiones, y reducir el sistema inicial  $J$  dimensional a uno  $A$  dimensional, con  $A < J$  manteniendo la información relevante del sistema. Geométricamente es un cambio de ejes, representando esas muestras en un sistema de coordenadas con un número inferior de ejes al utilizado inicialmente. Otra propiedad muy importante de los PCs es la ortogonalidad. El método busca las direcciones ortogonales que expliquen la máxima variabilidad de los datos de las muestras, utilizando estas direcciones como nuevos ejes de coordenadas o PCs.

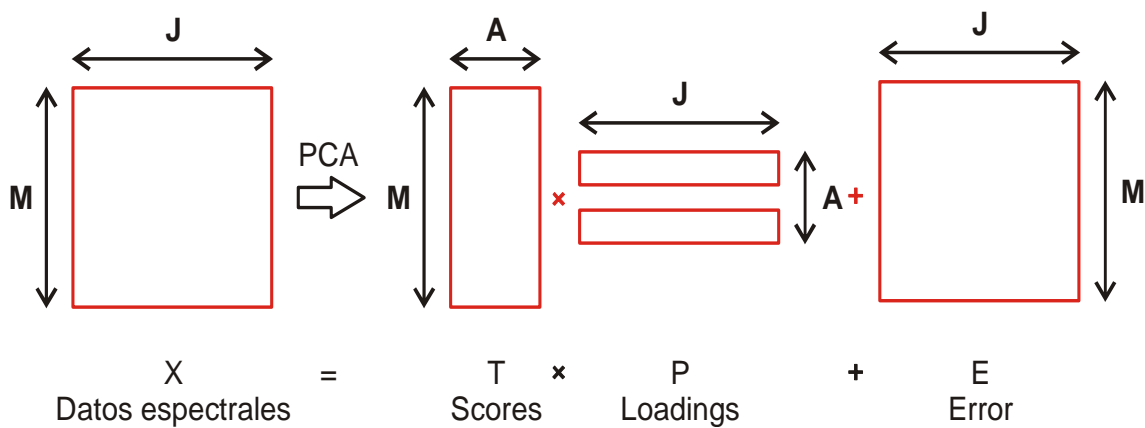


Figura 26. Descomposición de un conjunto de espectros mediante Análisis de Componentes Principales.

Resumiendo, PCA se utiliza para reducir el número de variables originales correlacionadas, a un número de variables no correlacionadas en un nuevo sistema de ejes ortogonales denominados PCs.

#### 4.2.4. Regresión Múltiple

La regresión es un análisis estadístico que se utiliza cuando el objetivo es predecir o explicar el valor de una variable a partir de los valores de otras.

La Regresión Múltiple permite valorar la relación (de forma predictiva) entre una variable dependiente y otras variables independientes, que pueden estar o no correlacionadas. El propósito general de la regresión múltiple es estudiar la dependencia de una variable criterio o respuesta cuantitativa respecto a un conjunto de otras variables predictoras o explicativas, valoradas también



---

cuantitativamente. Permite la predicción de valores para la variable dependiente en cada muestra, a partir de la combinación lineal de variables seleccionadas. El objetivo es encontrar los coeficientes de esta combinación que minimizan la suma de cuadrados residuales entre el valor predicho y el observado. El resultado de la aplicación de estas técnicas es una ecuación que representa la mejor recta de ajuste (la máxima correlación posible) entre las variables estudiadas, de la siguiente forma general:

$$Y = B_1V_1 + B_2V_2 + \dots + B_nV_n + \varepsilon \quad \text{Ec. 34}$$

Siendo:

$Y$  = valor predictivo de la variable dependiente

$B$  = coeficientes de regresión de cada una de las variables independientes

$V$  = variables independientes

$\varepsilon$  = constante que representa la predicción cuando las variables independientes son cero





## ***RESULTADOS Y DISCUSIÓN***



---

## CAPÍTULO 1

### METODOLOGÍA PARA LA ADQUISICIÓN DE IMÁGENES DIGITALES DE LAS UVAS Y SUS SEMILLAS. ESTUDIO DE LA EVOLUCIÓN COLORIMÉTRICA A LO LARGO DE LA MADURACIÓN

Francisco J. Rodríguez-Pulido, Luis Gómez-Robledo, Manuel Melgosa, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia, (2012). *Ripeness estimation of grape berries and seeds by image analysis*, Computers and Electronics in Agriculture, 82, 128-133, ISSN 0168-1699.

#### 1.1. Antecedentes

El color de los alimentos está estrechamente relacionado con su composición química, y es posible evaluarlo mediante la colorimetría triestímulo. La tridimensionalidad del color hace que podamos descomponerlo en tres atributos diferentes: claridad, que es una medida relativa de la luz reflejada frente a la absorbida y puede tomar valores entre 0 y 100; croma, que determina para cada tono o matiz la diferencia entre dicho color y el gris de la misma claridad y puede tomar valores positivos desde cero; y tono, que es el atributo de índole cualitativa, el cual hace que podamos clasificar un color como rojo, amarillo, etc. Estos atributos se expresan como  $L^*$ ,  $C_{ab}^*$  y  $h_{ab}$ , respectivamente, según el espacio de color CIELAB.

La evaluación del color se puede realizar haciendo uso de diferentes instrumentos, como colorímetros, espectrofotómetros o espectrorradiómetros. Sin embargo, estos instrumentos requieren de cierta uniformidad cromática en la muestra para que la medida del color sea representativa y, en ocasiones, esto es una tarea complicada y tediosa en la medida de muestras pequeñas, como son las semillas y las bayas de uva. En estos casos, es muy recomendable el uso de imágenes digitales para la medida del color. Además, mediante el uso del análisis de imagen no sólo se puede determinar el color, sino también otras características relacionadas con la morfología y la apariencia (Savakar & Anami, 2009; Zheng and Sun, 2008).

El propósito de la visión artificial es enseñar a un ordenador a que “entienda” las características de una escena. De este modo, en el caso de semillas y uvas, se pueden identificar en imágenes para, después, medir sus características morfológicas y colorimétricas, por ejemplo con fines predictivos o para la evaluación de la calidad.

Un equipo de visión artificial está formado por un sistema de iluminación, un dispositivo que captura las imágenes y convierte la señal analógica en digital (cámara digital), y un ordenador con el software apropiado para el procesado de las imágenes y la interpretación de los resultados (Wang & Sun, 2002). En este proceso, la cámara digital recibe las imágenes en su dispositivo CCD, formado por condensadores, que son estimulados por la radiación visible y registran el color en gradaciones de tres colores básicos: rojo, verde y azul (RGB), en concordancia con el carácter tridimensional del color. De todos los componentes que forman los equipos de visión artificial, la iluminación cobra una importancia crucial para la identificación de los objetos en la imagen, sobre todo por la estandarización de la luz que requiere el proceso del cálculo de color (CIE, 2007). RGB es un espacio de color aditivo diseñado para el registro y la reproducción del color en sistemas informáticos. Se utiliza para equipos provistos de pantallas y detectores de color, como televisores, cámaras digitales y teléfonos móviles, y su apariencia depende del dispositivo donde se muestre (Yam & Papadakis, 2004). En otras palabras, una imagen con las mismas coordenadas RGB se verá diferente, tendrá distinto color, según la pantalla donde se visualice. Por ello, la CIE recomienda el espacio de color CIELAB que, al ser independiente del dispositivo donde se muestre, elimina el problema de la reproducción del color que presenta el espacio RGB. El espacio de color CIELAB se suele utilizar en las medidas colorimétricas de alimentos (Brosnan & Sun, 2004; González-Miret et al., 2007), y la conversión entre espacios RGB y CIELAB requiere un proceso de calibrado con materiales de referencia (León et al., 2006).

Por otro lado, la apariencia de las semillas cambia durante la maduración como consecuencia de los cambios químicos que sufren. Durante la maduración, los compuestos fenólicos que se encuentran en las capas más externas de las

semillas, polimerizan, se oxidan y equilibran la agresividad que las semillas inmaduras aportan. Si bien un análisis químico no puede ser sustituido por la evaluación visual, la estimación objetiva y automática del pardeamiento de las semillas puede orientar sobre su estado.

También es importante el seguimiento de las bayas durante la maduración para la detección temprana de problemas y la toma de decisiones agronómicas. Uno de los parámetros que se evalúan en las bayas es el tamaño, influenciado por el crecimiento natural de la baya y por la climatología (especialmente el régimen de lluvias), que afecta a la concentración de azúcares del mosto. El procedimiento usual para su evaluación es tomar 25 bayas al azar, medir con un calibre, y establecer el diámetro promedio. En este proceso se pueden cometer dos tipos de errores. El primero es la consideración esférica de las bayas y por ello se asume como válida una sola medida por uva. Además, si la forma de la uva se aleja de la esfericidad claramente, es el prejuicio del analista el que decide dónde se realiza la medida. El segundo error que se puede cometer viene provocado por la falta de rigidez de la uva, que suele deformarse levemente al aplicar el calibre sobre su superficie, lo que conlleva medidas erróneas. Durante la maduración, también se determina el porcentaje de envero como parámetro de evaluación. Aunque en el momento de la vendimia lo normal es que todas las uvas lleven tiempo enveradas, es decir, con el color definitivo correspondiente a la madurez, su monitorización ayuda al viticultor a controlar el seguimiento de la maduración de la uva. Esta determinación también se realiza manualmente, tomando un conjunto de uvas y calculando el porcentaje, por conteo, de uvas enveradas. El método generalmente no utiliza un número significativo de uvas, y además, puede resultar inexacto en la evaluación de uvas que estén parcialmente enveradas.

## **1.2. Objetivos**

En este trabajo se ha puesto a punto la metodología para la adquisición de las imágenes digitales en semillas y bayas de uva. Además, se ha evaluado la evolución de algunos parámetros estudiados durante la maduración y se han

estudiado según variedades. Los objetivos concretos planteados en este trabajo fueron:

- Adquisición de imágenes con el equipo DigiEye, utilizando diferentes configuraciones de fondo, entorno, disposición de los objetos de la imagen y determinación de las condiciones óptimas de medida.
- Exploración de las imágenes en el espacio de color RGB para establecer los criterios de segmentación, así como los algoritmos para la extracción de información colorimétrica y morfológica de las muestras bajo lenguaje de programación MATLAB.
- Estudio de la evolución colorimétrica y morfológica de las semillas de uva de diferentes variedades, y aplicación de técnicas estadísticas para la discriminación varietal a partir de las imágenes digitales.

### **1.3. Parte experimental**

Se muestrearon uvas de la Denominación de Origen “Condado de Huelva” durante la maduración de las cosechas de 2009 y 2010, de las variedades tintas Tempranillo y Syrah, y la variedad blanca Zalema. De esta última, autóctona en el Condado de Huelva, se tomaron muestras cultivadas en dos tipos de suelo, arena y barro. De una muestra bruta de dos kilos se tomaron al azar cien bayas. El tratamiento de muestra para las uvas se limitó a la limpieza de los restos de polvo antes de adquirir las imágenes. Por su parte, las semillas fueron extraídas de las uvas con la ayuda de un cuchillo romo y se eliminaron los restos de pulpa.

La *obtención de imágenes*, en ambos casos, se realizó con el sistema DigiEye® (Luo et al., 2001), evitando el contacto entre los objetos (uvas o semillas) para facilitar la segmentación. Asimismo, para optimizar el proceso de segmentación, se probaron diferentes superficies como fondo para la adquisición de las imágenes. De entre todas, se observó que una superficie blanca incrementaba el contraste entre las regiones de interés y el fondo. Se eligió una hoja de papel gruesa y blanca, que está considerada como una superficie



---

lambertiana (superficie cuya reflectancia no varía con el ángulo de visión) (Jaglarlz et al., 2006). Las imágenes se adquirieron en formato TIF de dos maneras diferentes: una de ellas con información de color RGB con una profundidad de 8 bits por canal, y otra con coordenadas  $L^*a^*b^*$  con una profundidad de 16 bits por canal. Las imágenes en RGB se utilizaron solamente para el proceso de segmentación. El *procesado de las imágenes y el tratamiento estadístico* se realizó con los programas MATLAB® (The MathWorks Inc., Natick, USA), Statistica® (StatSoft Inc., Tulsa, USA) y DigiFood® (Heredia et al., 2006).

El siguiente paso fue la *calibración espacial* de las imágenes obtenidas con DigiEye®. Para ese fin, se adquirieron varias imágenes del calibre con el que se midieron las uvas en las direcciones vertical y horizontal de la imagen, y se midió la distancia en píxeles que correspondían a 10 cm en el calibre. Al considerar que la diferencia de altura de la muestra podría afectar a la calibración espacial (por su diferencia de proximidad a la cámara), se realizó este ensayo a las diferentes alturas que tenían semillas y uvas, y en diferentes posiciones en la bandeja de muestras. En las medidas realizadas no se encontraron diferencias significativas ( $p < 0.05$ ) entre la relación milímetros y píxeles, y se estableció, finalmente, una relación de 11.061465001 mm/píxel para todas las imágenes.

La *segmentación de las semillas* se realizó en el espacio de color HSI (del inglés *hue, saturation, intensity*). Éstos atributos de color se obtienen mediante transformaciones matemáticas directas a partir de RGB, y su ventaja es que tienen el mismo significado que los atributos tono, croma y claridad, respectivamente. A partir de los histogramas de las semillas de uva en todos los estadios de maduración, se aplicó un criterio de segmentación basado en umbralización. Buscando los valores mínimos en los histogramas entre el fondo y la región de interés, se seleccionaron todos los píxeles con valores de  $S$  entre 0 y 180 y valores de  $I$  entre 0 y 180. El criterio de segmentación para el reconocimiento de las uvas seleccionó todos los píxeles cuyo valor de  $I$  estuviera entre 0 y 160. Algunas de las imágenes contenían manchas de pulpa, además de píxeles aislados con valores comprendidos dentro de los límites del criterio de segmentación. Por ello se aplicó, además, un criterio de segmentación morfológico. En las imágenes de uvas se

eliminaron los objetos identificados cuyo diámetro medio de Feret no estuviera entre 5 y 25 mm, magnitudes que, con un amplio margen, permitían la identificación de las uvas eliminando los problemas en la segmentación. De la misma manera, en las imágenes de semillas sólo se consideraron los objetos con un área entre 6 y 35 mm<sup>2</sup> y cuya longitud fuera menor de 15 mm.

En los dos sets de imágenes se midieron las *coordenadas colorimétricas* CIELAB de todos los píxeles de la región de interés y las *características morfológicas* de todos los objetos. Una vez extraídos estos datos, se evaluó su evolución a lo largo de la maduración. En cuanto a las semillas, se observó que los datos morfológicos difirieron entre las variedades de uva estudiadas aunque, dentro de cada variedad, se mantuvieron prácticamente estables durante todo el periodo. Sin embargo sí se halló evolución en los valores de los parámetros de color (algo esperable ya que incluso eran apreciables *de visu*). Es por ello que se utilizaron los datos de morfología para evaluar las diferencias varietales, y los datos colorimétricos para evaluar el estado de madurez.

En el *tratamiento estadístico* de los datos, y utilizando como variables todas las características morfológicas, se aplicó un análisis de la varianza (ANOVA). Se encontraron diferencias significativas ( $p < 0.05$ ) entre la morfología de las semillas de uva según variedad, aunque no se encontraron diferencias significativas entre los dos tipos de suelo para la variedad Zalema. También se aplicó un análisis discriminante por pasos (*Forward Stepwise Discriminant Analysis*), el cual clasifica cada semilla dentro de una variedad a partir de las variables seleccionadas. Este método incluye iterativamente variables hasta que la inclusión de más variables deja de optimizar el modelo. Los resultados de la clasificación obtenida en el análisis discriminante están resumidos en la Tabla 2. Al igual que en el ANOVA, el análisis discriminante no pudo diferenciar entre los dos tipos de suelo en la variedad Zalema. De las variables utilizadas para la clasificación, el análisis discriminante incluyó secuencialmente *relación de aspecto*, *grado de redondez* y *ancho* de semilla en las ecuaciones discriminantes como las de mayor influencia en la clasificación.

Tabla 2. Resumen de los resultados del análisis discriminante. Clasificación (a) según semillas de uva blanca o tinta, (b) según variedad, y (c) según variedad incluyendo tipo de suelo.

Clasificación				
		(a)	(b)	(c)
tinta	Syrah	100%	100%	100%
	Tempranillo		87.7%	87.7%
blanca	Zalema arena	100%	100%	57.2%
	Zalema barro			57.2%

En cuanto a la evolución de las coordenadas colorimétricas en las semillas,  $L^*$ ,  $C_{ab}^*$  y  $h_{ab}$  disminuyeron durante la maduración, al principio rápidamente y luego más lento, hasta estabilizarse en los últimos estadios. Además, el patrón que describió la evolución del color fue diferente según variedad de uva blanca o tinta.

En la primera cosecha de las consideradas en este estudio (2009), se observó que el pardeamiento de la semilla no ocurría de manera homogénea sino que, durante la maduración, aparecían zonas oscuras en la parte más ancha de la semilla y se extendían poco a poco hasta completar la semilla entera. Con objeto de evaluar este pardeamiento, se examinaron los histogramas pertenecientes a muestras de semillas antes y después de este fenómeno. La superposición de ambos histogramas presentaba un valle en el valor medio de claridad ( $L^*=50$ ). Así, se propuso la definición de un “índice de pardeamiento” como la fracción de píxeles cuyo valor de  $L^*$  es menor a 50 unidades CIELAB. Este índice aumentó rápidamente en los primeros estadios de madurez y se estabilizó en los últimos muestreos. Para probar la capacidad del índice propuesto, se evaluaron las semillas de la siguiente campaña (2010) y se observó que de nuevo los histogramas de claridad de las muestras inicial y final se cruzaban en el mismo valor de  $L^*=50$ . La evolución también fue similar a la de la campaña anterior, confirmando la capacidad del parámetro  $L^*$  para ser usado para establecer el “índice de pardeamiento”.

Otro de los aspectos que llamó la atención fue la distribución de los puntos de color pertenecientes a todos los píxeles de las semillas de uva en el espacio CIELAB (Figura 27). Estas nubes de puntos se desplazaron en el diagrama ( $a^*b^*$ ) modificando el color medio, y también cambiaron de forma. Se observó que la extensión y la forma de esas nubes de puntos podían reflejar la heterogeneidad cromática de las correspondientes muestras. Sin embargo, en esa configuración no

era posible cuantificar la heterogeneidad en términos de esas nubes de puntos. Con el programa Statistica® se realizó un ajuste de los puntos que conforman la nube a una elipse, sin tener en cuenta la coordenada cromática claridad (por tanto, bidimensional), y se encontraron interesantes correlaciones entre las dimensiones de las elipses y otras variables colorimétricas, por ejemplo con el croma.

El resultado de este estudio preliminar fue el desarrollo de una metodología para la cuantificación de estas nubes de puntos en términos de heterogeneidad, ajustándolas a elipsoides de color y determinando el número de colores discernibles que contiene cada nube. Esta metodología se desarrolla en profundidad en el Capítulo 2 de esta Memoria.

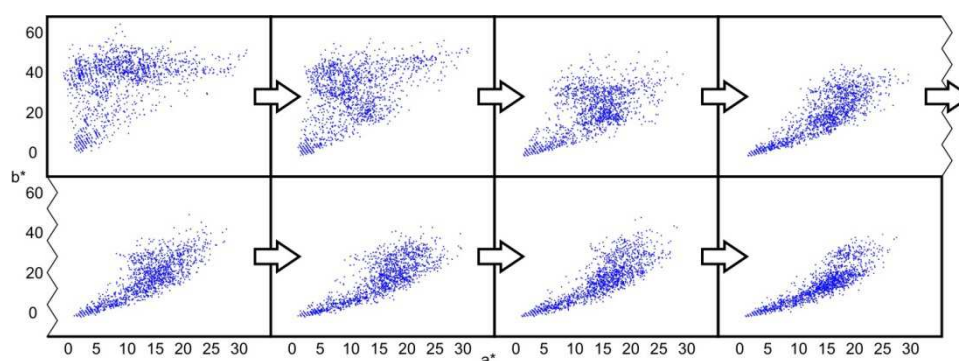


Figura 27. Evolución de las nubes de puntos de las semillas de uva en el diagrama  $a^*b^*$  durante la maduración.

### 1.4. Conclusiones

- Se ha establecido la metodología de adquisición de imágenes digitales de semillas de uva para la determinación de sus características colorimétricas y morfológicas. De la misma manera, se ha comprobado la aplicabilidad de la metodología propuesta para evaluar de manera objetiva algunos parámetros de interés de la maduración de las bayas de uva, como son el nivel de envero y el calibre.
- Se ha evaluado la evolución de estas características durante los dos últimos meses de maduración antes de la vendimia, en dos campañas consecutivas. No se observó evolución en la morfología de las semillas dentro de cada

variedad de uva, pero se encontraron diferencias entre variedades, lo que justifica el uso de estas variables para la discriminación varietal. Entre estas variables, la relación de aspecto y el grado de redondez tuvieron mayor peso en las funciones discriminantes.

- Se ha propuesto un índice de pardeamiento, basado en coordenadas colorimétricas, para evaluar, de manera objetiva e instrumental, la apariencia de las semillas. La idoneidad de este índice ha sido constatada en muestras de diferentes variedades y en dos cosechas.
- En general, se ha puesto de manifiesto la utilidad de la visión artificial y la colorimetría triestímulo para evaluar objetivamente las propiedades de apariencia y semillas de uva. Los programas desarrollados bajo lenguaje MATLAB® para la aplicación de esta metodología han podido ser inscritos en el Registro de la Propiedad Intelectual.





## Ripeness estimation of grape berries and seeds by image analysis

Francisco J. Rodríguez-Pulido<sup>a</sup>, Luis Gómez-Robledo<sup>b</sup>, Manuel Melgosa<sup>b</sup>, Belén Gordillo<sup>a</sup>,  
M. Lourdes González-Miret<sup>a</sup>, Francisco J. Heredia<sup>a,\*</sup>

<sup>a</sup> Food Colour & Quality Lab., Department of Nutrition & Food Science, Facultad de Farmacia, Universidad de Sevilla, 41012-Sevilla, Spain

<sup>b</sup> Department of Optics, Faculty of Sciences (Mecenas Building), University of Granada, 18071-Granada, Spain

### ARTICLE INFO

#### Article history:

Received 23 July 2011

Received in revised form 19 November 2011

Accepted 6 January 2012

#### Keywords:

Colour

Grape seeds

Image analysis

Ripeness

*Vitis vinifera*

### ABSTRACT

Digital imaging has become a powerful tool for the characterization and quality control of foodstuff. Because of the need to automate processes, faster tools are needed and Computer Vision is a good alternative to chemical analysis of many products in quality control. Appearance of grape seeds and grape berries change during the ripeness. These changes are closely related to the chemical composition, especially phenolics, which are very important compounds due to their implications on the intensity and stability of red wine colour. In this study, a complete characterization of grape seeds and grape berries by digital image analysis is described. The size of grapes and the veraison has been determined by image analysis and it has been also established an objective Browning Index of seeds. Morphological differences between varieties were studied by applying discriminant analysis models which allowed us to classify the grape seeds with high accuracy.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

### 1.1. Colorimetry and food stuff

Colour and appearance are closely related to sensory properties and chemical composition of food. Normally, colour is measured by tristimulus colorimetry. The colour stimulus is composed of three different sensations, giving to colour three-dimensional nature. These attributes are:

- Lightness: This feature makes a colour lighter or darker. It is a relative measure of the reflected light against the absorbed. Value 0 is assigned to black and value 100 is assigned to white.
- Chroma: It determines for each hue, the colour difference from the grey having the same lightness. It can take positive values from zero.
- Hue: It is the main attribute. It is a qualitative property which allows classifying colours as red, yellow, etc. It is related to differences in absorbance of radiant energy at different wavelengths. Hue is specified as an angle.

These attributes are often expressed as  $L^*$ ,  $C_{ab}^*$  and  $h_{ab}$ , respectively, according to the CIELAB colour space.

It can be used different kinds of instruments, such as colorimeters, spectrophotometers and spectroradiometers. Nevertheless,

these instruments require homogenizing the sample to achieve uniform colour, which becomes tedious and complicated task to measure colour of heterogeneous stuff or small objects, such as grape berries and grape seeds. In these cases, it is very advantageous the use of digital images for this purpose. Digital image analysis appears as successful complement since it can be determined not only colour but also other characteristics such as shape, texture and homogeneity (Savakar and Anami, 2009; Zheng and Sun, 2008b).

### 1.2. Imaging hardware and colour spaces

Computer Vision is a subfield of Artificial Intelligence. The aim of Computer Vision is 'teach' a computer for understanding a scene or the characteristics of an image. In this way, we can identify the seeds and grapes in these images, and then, morphological and colorimetric characteristics can be extracted from each one. Computer Vision is a powerful tool for testing quality in alimentary industry (Brosnan and Sun, 2004).

A Computer Vision system includes: an illumination system, a charge-coupled device (CCD), a frame grabber which converts the analogue image from the camera into a digitized one, and a computer with the suitable software for image processing and interpretation of results (Wang and Sun, 2002). Illumination system has high importance, not only because of the need to identify objects in the image, but because the colour calculation process needs a light standardization (CIE, 2007). Once the illumination is controlled, a digital camera receives images onto a CCD. It has

\* Corresponding author. Tel.: +34 954556495; fax: +34 954557017.

E-mail address: [heredia@us.es](mailto:heredia@us.es) (F.J. Heredia).

capacitors which are stimulated by visible radiation and it is registered in gradations of three basic colours: red, green and blue (RGB). This is consistent with the theory that every colour can be reproduced by the combination of three primary colours.

The RGB colour space is an additive colour model that uses transmitted light to display colours. It is used for television and other devices screens, so this model is device-dependent (its appearance depends on the display) (Yam and Papadakis, 2004). The  $L^*a^*b^*$  model is an international standard for colour measurement developed by the *Commission Internationale de l'Eclairage* (CIE) in 1976. Here,  $L^*$  has the same meaning that the  $L^*$  described previously.  $a^*$  (from green to red) and  $b^*$  (from blue to yellow) are Cartesian coordinates of Polar coordinates  $C_{ab}^*$  and  $h_{ab}$  previously described. This colour space is device-independent, providing consistent colour regardless of the input or output device such as digital camera, scanner, monitor and printer. CIELAB values are frequently used in food research (González-Miret et al., 2007).

Due to RGB colour space is not continuous each channel can take only integer values between 0 and 255. In order to calculate the colorimetric coordinated recommended by *Commission Internationale de l'Eclairage* (CIE), it is necessary to transform from RGB to CIELAB colour spaces. This transformation requires calibration and it depends on illumination when images are taken (León et al., 2006).

One of the most important steps in image analysis is segmentation. Segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something more meaningful and easier to analyze (Zheng and Sun, 2008a).

### 1.3. Advantages of using image analysis for grapes and grape seeds characterization

The size of the grapes is measured in routine analysis during ripening. The diameter is normally measured with a calliper, which measures the size in only one direction and the mean of several grapes is given, making this a tedious task. In addition, the berry could be accidentally tightened during measurement and it induces error. For this purpose, the use of Feret diameter has been proposed (González Marcos et al., 2006). Feret diameter is the maximum length of chord for a counterclockwise angle with the  $x$  axis, defined between  $0^\circ$  and  $180^\circ$  (Fig. 1). Due to grape berries are not completely spherical, for each berry, the mean of all Feret diameters for all possible angles have been considered. This way, the size will be given as an average of all possible Feret diameters of all the grapes analyzed.

During the grapes ripening there is a loss of chlorophyll and the formation of the final dyes. This process starts at the veraison, the onset of ripening, when the colour of the grape berries change, being the transition from berry growth to berry ripening. In red

grapes, the colour changes from green to purple, even almost black. The veraison is normally expressed as a percentage determined by visual inspection. It is possible to establish an objective veraison index by image analysis.

There are more appearance-related characteristics of the seeds and grapes associated to maturity, such as the browning of the seeds (Ristic and Iland, 2005). All these parameters can be estimated of an objective and automated manner by image analysis.

## 2. Material and methods

### 2.1. Sampling

The vineyards sampled are included under the “Condado de Huelva” Designation of Origin, in southwestern Spain, harvested in 2009 and 2010. Two red varieties (*Tempranillo* and *Syrah*) and one autochthonous white variety (*Zalema*) cultivated in two kinds of soil (*Sand* and *Clay*) were used. Samples were taken twice a week from early July until harvest (which occurred approximately at end of August depending on variety). Sampling was carried out taking a pair of berries from alternate grapevines and from both sides up to reach 2 kg of berries, ensuring this way the representativeness of the sample. Once in lab, one hundred berries were randomly taken and cleaned for acquiring the images. Seeds of these berries were removed and dried at room temperature for two hours before acquiring the image.

### 2.2. Apparatus

The DigiEye® imaging system based upon the calibrated digital camera was used (Luo et al., 2001). It includes an illumination box specially designed by VeriVide Ltd. (Leicester, UK) to illuminate the samples consistently and a digital camera connected to a computer (Fig. 2).

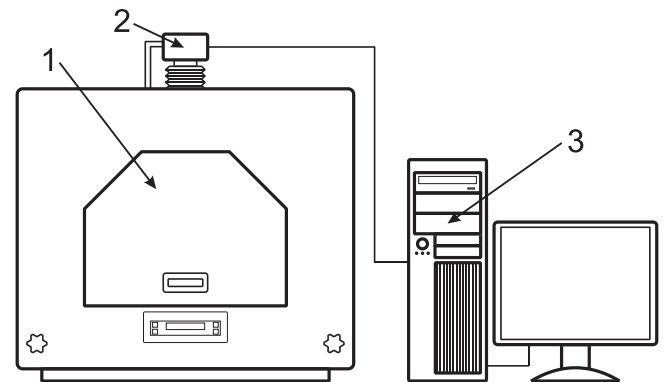


Fig. 2. The DigiEye® System: (1) illumination box; (2) digital camera; (3) computer.

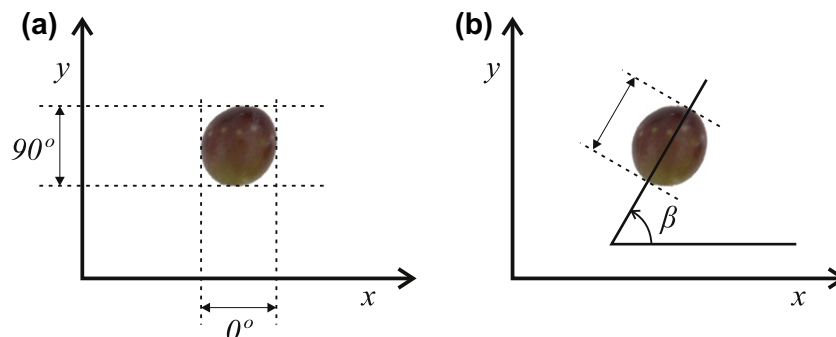


Fig. 1. (a) Feret diameters at  $0^\circ$  and  $90^\circ$ . (b) Feret diameter at an angle  $\beta$ .



The digital camera used for image acquisition was 10.2-mega-pixel Nikon® D80 with Nikkor® 35 mm f/2D objective. The camera was connected via USB to a computer with Pentium IV processor at 3.00 GHz. The cabinet is equipped with two fluorescent tubes that emulate the standard illuminant D65 and offer stable lighting conditions. Lamps were switched on at least ten minutes before being used, according to manufacturer indications, to stabilize them.

For obtaining morphological and appearance parameters, as well as CIELAB coordinates from RGB colour space, the software DigiFood® (Heredia et al., 2006) was used.

### 3. Results

#### 3.1. Image acquisition

The camera parameters were set up for all images as shown in Table 1. To ensure that the camera always takes images the same way, calibration is required prior to use by a certified colour chart DigiTizer provided by VeriVide Ltd. (Leicester, UK) (Fig. 3), composed of a matrix of  $12 \times 20$  squares of different colours. DigiEye have data of the  $L^*a^*b^*$  coordinates of each square and adapt the RGB coordinates obtained by the camera for this calibration.

Several backgrounds were tested for image acquisition of grapes and seeds. Among them, it was found that the white surface increases the contrast between objects and background, and improves the segmentation process. Once the background was selected, both seeds and berries were put avoiding contact among them over a white thick sheet, considered as a good Lambertian surface (diffuse reflectance surface which does not vary depending on the viewing angle) (Jaglarlz et al., 2006).

#### 3.2. Image processing

Although the analytical information was obtained from the CIELAB coordinates calculated, HSI (hue, saturation and intensity) colour space was used for segmentation process (Jack, 2008). HSI

colour space is easier to understand than RGB, due to its likeness with the  $L^*$ ,  $C_{ab}^*$  and  $h_{ab}$  stimuli.

The histogram thresholding segmentation technique was used (Cheng et al., 2001). It allows distinguish different objects in images that are composed of regions with different colour ranges. Histograms of HIS values were plotted, containing well separated peaks corresponding to the sample and the background. The lowest point between adjacent peaks was stated as threshold for segmentation process. After establishing the thresholds, the suitability of the criteria was confirmed by visual inspection.

For grape berries recognition, all regions having intensity (I) between 0 and 160 units were considered. Restrictions in hue (H) and saturation (S) values were not required for grapes berries. To prevent the recognition of small spots as grapes, the segmentation was restricted to objects having Feret diameter between 5 and 25 mm.

In the same way, for seeds recognition, sets of pixels with saturation (S) between 0 and 180 units and intensity (I) between 0 and 140 units were considered. All hues (H) were considered. Similarly, morphological restrictions were applied. All objects must have area between 6 and 35 mm<sup>2</sup> and length shorter than 15 mm for being considered (Fig. 4).

The following morphological and appearance parameters were measured from segmented images of seeds:

RGB colour was obtained from the raw images. CIELAB colour coordinates were obtained from the RGB values and through the camera calibration. All images were acquired under the same conditions. Thus, taking a picture of the calliper, it was possible to calculate the conversion between units of length and pixels. The size of seeds was measured in two directions: diameter along major axis of the seed (length) and diameter along the axis perpendicular to the major axis (width). In order to measure the elongation degree of each seed, aspect ratio was calculated, which is ratio is the ratio between major axis and minor axis of the ellipse equivalent to the seed. Roundness values were calculated using the following dimensionless equation (González Marcos et al., 2006):

$$\text{Roundness} = \frac{\text{perimeter}^2}{4 \times \pi \times \text{area}}$$

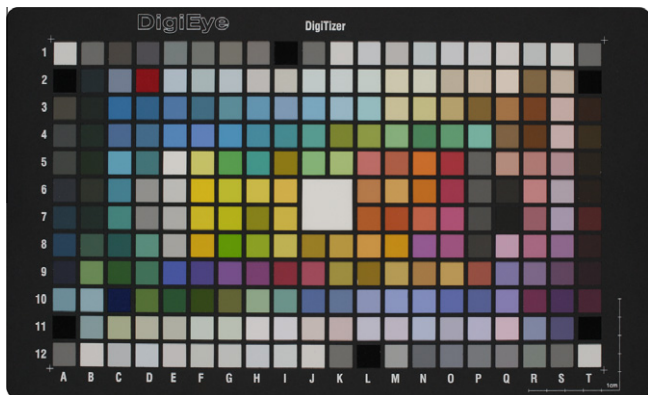
In this sense, a perfect circle has a roundness value of 1 and it becomes higher while the shape elongates. Some authors use the inverse of this equation for the same purpose. The area of each pixel was calculated by the spatial calibration. Thus, the area of each seed was calculated by pixel count.

Measuring colour of each pixel, it is possible to have an estimation of the heterogeneity of the sample as the fraction of pixels that

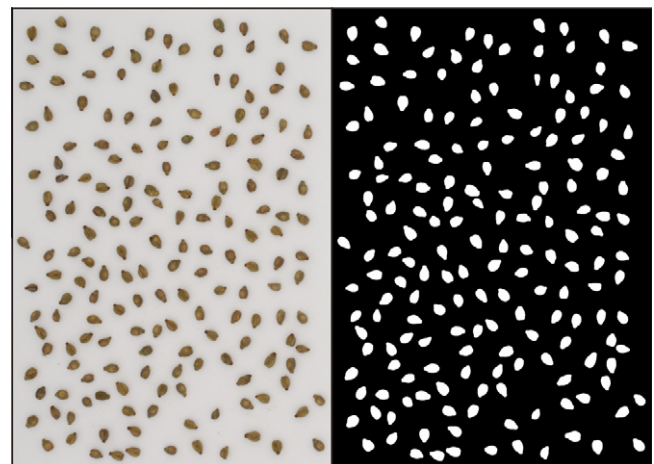
**Table 1**

Camera setup.

Time exposure	1/15 s
Aperture	f/6.3
Sensitivity	200 ISO
Format	TIF
Image size	3872 × 2592
Resolution	96 pixels per inch (ppi)



**Fig. 3.** Colour chart used for the calibration of the digital camera. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Image of Syrah seeds and the resulting mask after segmentation.

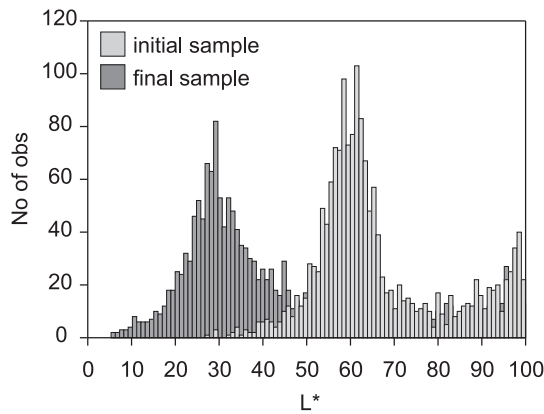
**Table 2**  
Some morphological parameters measured in seeds regarding the variety (mean ± Std. Dev.).

	Syrah	Tempranillo	Zalema (sand)	Zalema (clay)
Length (mm)	6.6 ± 0.2 <sup>a</sup>	6.8 ± 0.2 <sup>a</sup>	5.7 ± 0.1 <sup>b</sup>	5.7 ± 0.1 <sup>b</sup>
Width (mm)	3.8 ± 0.1 <sup>b</sup>	4.0 ± 0.1 <sup>a</sup>	3.6 ± 0.1 <sup>c</sup>	3.6 ± 0.1 <sup>c</sup>
Aspect ratio	1.80 ± 0.01 <sup>a</sup>	1.72 ± 0.01 <sup>b</sup>	1.64 ± 0.01 <sup>c</sup>	1.64 ± 0.01 <sup>c</sup>
Roundness	1.31 ± 0.02 <sup>a</sup>	1.26 ± 0.03 <sup>b</sup>	1.24 ± 0.03 <sup>b</sup>	1.24 ± 0.02 <sup>b</sup>
Perimeter (mm)	16.91 ± 0.08 <sup>b</sup>	17.50 ± 0.09 <sup>a</sup>	14.85 ± 0.07 <sup>c</sup>	14.76 ± 0.07 <sup>c</sup>
Area (mm <sup>2</sup> )	17.3 ± 0.7 <sup>b</sup>	19.1 ± 1.1 <sup>a</sup>	14.1 ± 0.9 <sup>c</sup>	13.9 ± 0.7 <sup>c</sup>
Heterogeneity (%)	17.5 ± 0.6 <sup>a</sup>	13.9 ± 0.7 <sup>b</sup>	15.4 ± 0.5 <sup>ab</sup>	14.7 ± 0.5 <sup>ab</sup>

<sup>a-c</sup> Different superscripts within rows for each parameter indicate statistically significant differences ( $P < 0.05$ ).

deviate more than 10% from the average intensity. Other indicator of heterogeneity is obtained by using the Mean Colour Difference from the Mean (MCDM), proposed by Berns (2000):

$$MCDM = \frac{\sum_{i=1}^N \left[ (L_i^* - \bar{L}^*)^2 + (a_i^* - \bar{a}^*)^2 + (b_i^* - \bar{b}^*)^2 \right]^{1/2}}{N}$$



**Fig. 5.** Overlap of the histograms of  $L^*$  for the initial and final samples.

### 3.3. Study of appearance and colorimetric evolution of grapes and grape seeds

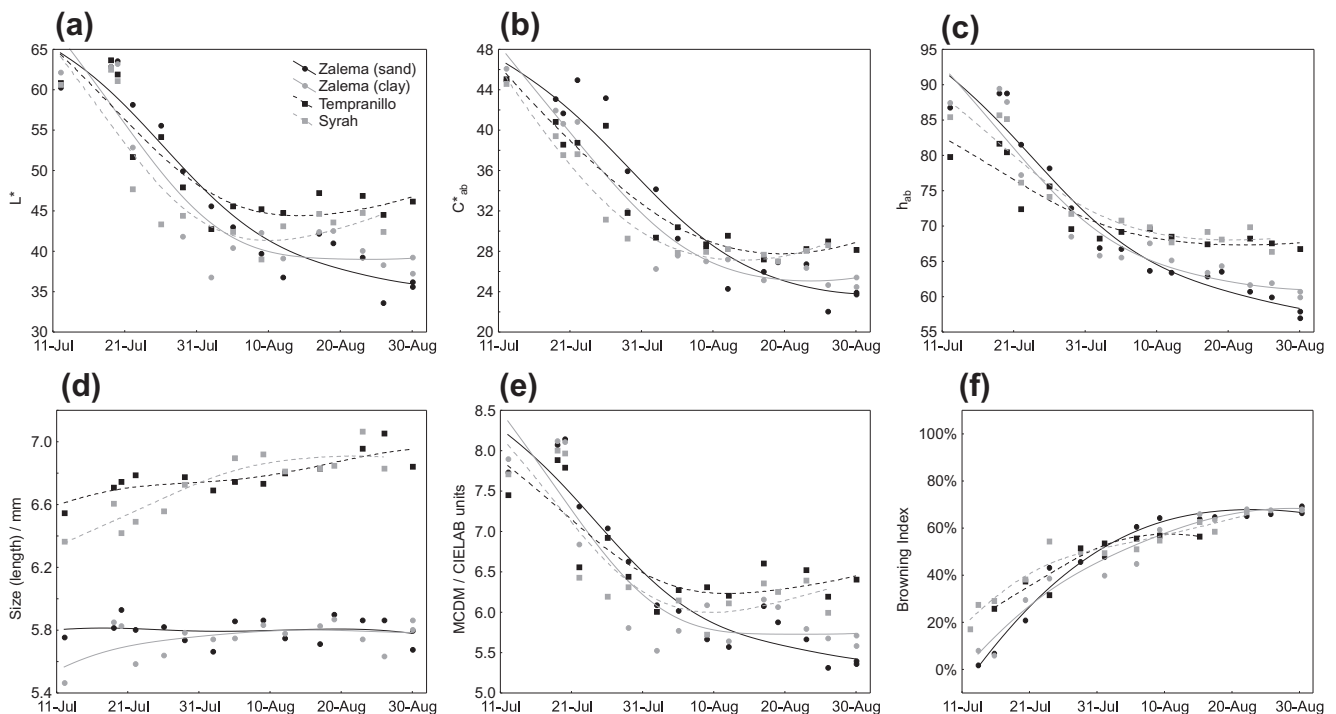
#### 3.3.1. Varietal classification by morphology

From morphological data of all samples corresponding to the four grapes varieties considered in this study, analysis of variance ANOVA with Statistica<sup>®</sup> 8.0 (StatSoft Inc., 2007) was applied. The aim of considering all samples was to distinguish among varieties regardless the moment of sampling. Significant differences ( $p < 0.05$ ) in the morphology of seeds from different grape varieties were found. No differences regarding the soil type for the white variety were found (Table 2).

Discriminant analysis was applied to classify the samples based on the features of appearance of the seed. Based on seeds from red grapes or white grapes, the classification was 100%. The classification between the two red varieties was 100% for Syrah and 87.5% for Tempranillo. When only the white variety was considered, it was not possible to discriminate correctly between soil types by morphological features. The classification for Zalema Sand was 71.43% while for Zalema Clay was only 57.14%. In general, the parameters aspect ratio, roundness and width gain more weight in the equations.

#### 3.3.2. Evolution of the appearance of seeds during maturation

According to the variety, similar changes occur in grape seed during the maturation. Colorimetric parameters decrease quickly



**Fig. 6.** Evolution of the Browning Index of seeds during the ripening.

in the first stages and stabilize themselves in the last ones (Fig. 6a–c). However, a related fit is visible in the figure depending on the type of grape (red or white). There was no noteworthy change in the shape of the seeds in this period. Nevertheless, the graph

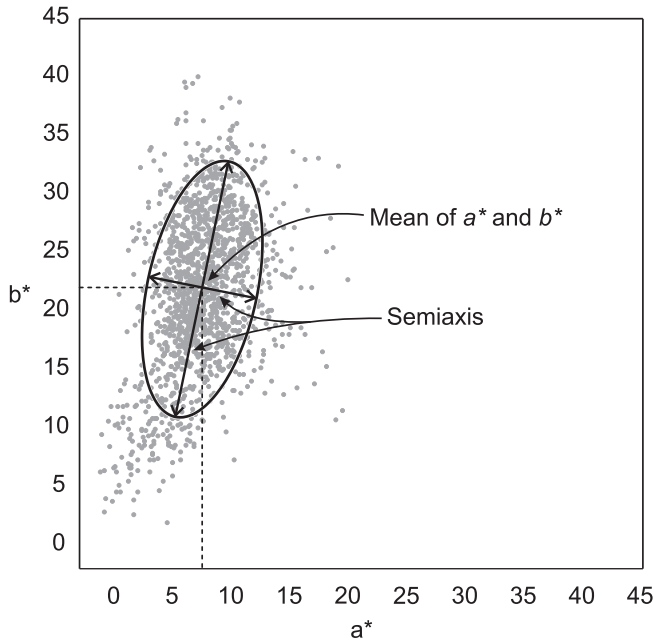


Fig. 7. Scatterplot of seed colour and ellipse fitting.

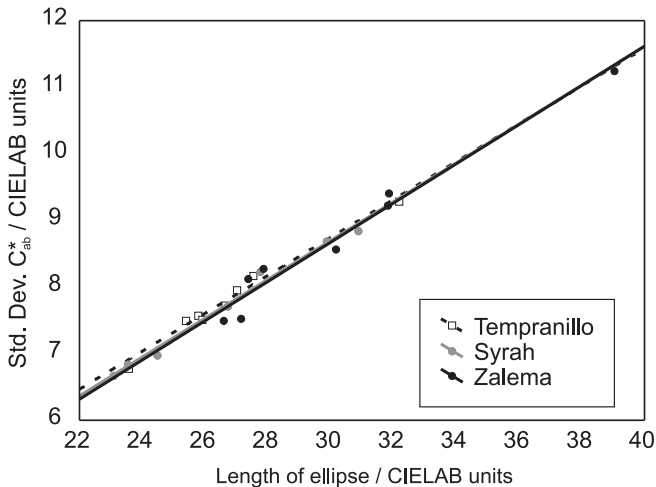


Fig. 8. Correlation between Std. Dev. of chroma ( $C_{ab}^*$ ) and length of ellipse.

clearly shows the difference between the size of the seeds of red and white varieties (Fig. 6d). These differences were essential in the discriminant analysis described above. The Fig. 6f shows how natural darkening of seeds produces the decrease of the mean CIELAB colour difference from the mean MCDM, being this latest colour more homogeneous.

**3.3.2.1. Browning Index.** Colour heterogeneity increases in the early stages of maturation and decreases in the final stages, because browning process does not occur in a homogeneous manner. The seeds become dark brown bit by bit while the light green areas disappear in the same way. Histograms of  $L^*$  for the initial and final samples were overlaid in order to study the behaviour. Both histograms crossed in the value  $L^* = 50$  (Fig. 5). Thus, “Browning Index” was defined as the fraction of pixel having  $L^*$  value lower than 50 CIELAB units.

$$\text{Browning Index} = \frac{\text{No. of pixels with } L^* < 50 \text{ units}}{\text{No. total pixels}}$$

During maturation, the Browning Index increased rapidly at the early stages while it is stabilized at the final stages when the seed was almost completely dark brown (Fig. 6f).

**3.3.2.2. Colour ellipses.** By spectroradiometry, we obtain the average from the measured area. In colour analysis by digitalization, we obtain a measure from each pixel from the image.

After processing the images, all colour points corresponding to a sample were plotted in the  $a^*b^*$ -diagram. Regarding the evolution of these “clouds of points” it can be appreciated how the chromatic dispersion decreases over time. In the early stages, the seeds are pale green mixed with light brown, which evolve until the harvest, producing a clear convergence towards dark brown colours. These points give information about the chromatic heterogeneity of the sample, and they can be fitted to an ellipse for its quantification. The ellipse is oriented along the line of least squares linear fit of these points. The centre of the ellipse coincides with the mean of  $a^*$  and  $b^*$  values and its semiaxis are proportional to the dispersion of the points (Fig. 7). It was found high correlation ( $r^2 = 0.98$ ) between the length of the ellipses containing the pixels of one sample and the standard deviation of chroma corresponding to these pixels ( $C_{ab}^*$ ) (Fig. 8).

**3.3.3. Evolution of the appearance of grapes during maturation**

A group of berries were taken and their veraison were assigned by visual assessment. The unripe berries are clearly green. Then, its colour turns lighter and begins to appear pink zones. Finally, the skin colour moves forward purple almost black. For each berry, the veraison is considered complete when the colour reaches this deep purple.

A criterion based on the bright was enough to assess the veraison level. After identifying the grape berries within the images, the

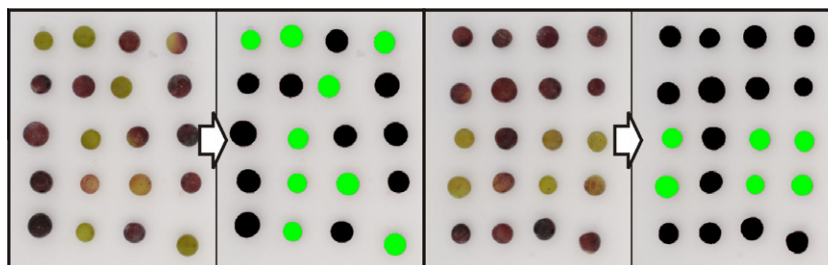


Fig. 9. Segmentation mask and veraison assessment by the selected criterion.

RGB values of grapes before and after veraison were studied. The following threshold was established:

$$\frac{R + G + B}{3} \geq 90 \rightarrow \text{pre-veraison,}$$

$$\frac{R + G + B}{3} < 90 \rightarrow \text{post-veraison.}$$

As performed by visual assessment, the veraison was calculated for each sample (group of grapes) as the percentage of berries that meet the threshold set (Fig. 9).

#### 4. Conclusions

The appearance of grape seeds has been assessed along two months corresponding to the final stages of ripening in two vintages. Within the same grape variety, the morphology of the seeds does not change during the period considered. However, there are significant differences ( $p < 0.05$ ) among varieties, so we can use these parameters for varietal discrimination. Except for the Zalema cv., which two kinds of samples were considered regarding the type of soil, classification was successful, being the aspect ratio and the roundness the variables that induced the differences.

The threshold chosen from the overlap of initial and final histograms of lightness ( $L^*$ ) was the same for the two vintages considered (2009 and 2010). So, the Browning Index proposed is a good criterion for estimating the maturity of grapes seeds.

The colour ellipses containing most of pixels from an image are measurable, clearly express the average colour and the chromatic heterogeneity of these samples, and there is a clear relationship between the size of the ellipse and the standard deviation of chroma ( $C_{ab}^*$ ).

Veraison and the size of the grape berries are characteristics routinely assessed during maturation. It has been established an effective and objective criteria for the veraison assessment as well as a measure of size faster and more accurate than that obtained manually using a calliper based on digital image analysis.

In short, the image analysis is becoming a fast and successful tool for the evaluation of food products. It is possible to use these tools to characterize the colour, morphology and appearance of the grapes and their seeds in a comprehensive manner.

#### Acknowledgements

We thank the Ministry of Science and Innovation of Spain (Project AGL2008-05569-C02-02) and the concession of the fellowship (BES-2009-025429) for financial support.

#### References

- Berns, R., 2000. Billmeyer and Saltzman's Principles of Color Technology. Wiley, New York.
- Brosnan, T., Sun, D.W., 2004. Improving quality inspection of food products by computer vision – a review. *Journal of Food Engineering* 61, 3–16.
- Cheng, H.D., Jiang, X.H., Sun, Y., Wang, J., 2001. Color image segmentation: advances and prospects. *Pattern Recognition* 34, 2259–2281.
- CIE, 2007. Commission internationale de l'Eclairage. Standard Illuminants for Colorimetry. ISO 11664-2:2007.
- González Marcos, A., Martínez de Pisón Ascacibar, F.J., Pernía Espinoza, A.V., Alba Elías, F., Castejón Limas, M., Ordieres Meré, J., Vergara González, E., 2006. Técnicas y Algoritmos Básicos de Visión Artificial. Universidad de La Rioja, Logroño.
- González-Miret, M.L., Ji, W., Luo, R., Hutchings, J., Heredia, F.J., 2007. Measuring colour appearance of red wines. *Food Quality and Preference* 18, 862–871.
- Heredia, F.J., Gonzalez-Miret, M.L., Álvarez, C., Ramírez, A., 2006. DigiFood. Registro N° SE-01298.
- Jack, K., 2008. Color spaces. In: *Digital Video and DSP*. Newnes, Burlington, pp. 15–29.
- Jaglariz, J., Duraj, R., Szopa, P., Cisowski, J., Czernastek, H., 2006. Investigation of white standards by means of bidirectional reflection distribution function and integrating sphere methods. *Optica Applicata* 36, 97–103.
- León, K., Mery, D., Pedreschi, F., León, J., 2006. Color measurement in  $L^*a^*b^*$  units from RGB digital images. *Food Research International* 39, 1084–1091.
- Luo, M.R., Cui, G.H., Li, C., 2001. British Patent (Application No. 0124683.4) entitled Apparatus and method for measuring colour (DigiEye System). Derby University Enterprises Limited.
- Ristic, R., Iland, P.G., 2005. Relationships between seed and berry development of *Vitis Vinifera* L. cv Shiraz: developmental changes in seed morphology and phenolic composition. *Australian Journal of Grape and Wine Research* 11, 43–58.
- Savakar, D.G., Anami, B.S., 2009. Recognition and classification of food grains, fruits and flowers using machine vision. *International Journal of Food Engineering* 5, Article 14.
- StatSoft Inc., 2007. Statistica 8.0. Tulsa, USA.
- Wang, H.H., Sun, D.W., 2002. Correlation between cheese meltability determined with a computer vision method and with Arnott and Schreiber tests. *Journal of Food Science* 67, 745–749.
- Yam, K.L., Papadakis, S.E., 2004. A simple digital imaging method for measuring and analyzing color of food surfaces. *Journal of Food Engineering* 61, 137–142.
- Zheng, C., Sun, D.W., 2008a. Image segmentation techniques. In: Da-Wen, S. (Ed.), *Computer Vision Technology for Food Quality Evaluation*. Academic Press, Amsterdam, pp. 37–56.
- Zheng, C., Sun, D.W., 2008b. Object measurement methods. In: Da-Wen, S. (Ed.), *Computer Vision Technology for Food Quality Evaluation*. Academic Press, Amsterdam, pp. 57–80.

---

## CAPÍTULO 2

### EVALUACIÓN DEL COLOR MEDIANTE TÉCNICAS DE ANÁLISIS DE IMAGEN Y ESTABLECIMIENTO DE LA METODOLOGÍA PARA LA CUANTIFICACIÓN DE NUBES DE PUNTOS

Francisco J. Rodríguez-Pulido, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia, (2013). *Analysis of food appearance properties by computer vision applying ellipsoids to colour data*, Computers and Electronics in Agriculture, 99, 108-115, ISSN 0168-1699.

#### 2.1. Antecedentes

La visión artificial ha cambiado el panorama de la evaluación de la composición de los alimentos en los últimos años. Los avances en estas técnicas junto al uso de potentes herramientas quimiométricas permiten que se pueda conocer la distribución espacial de la composición de una muestra. Además, el color y la apariencia son unas de las características más evaluadas en un alimento por su estrecha relación con la calidad y la preferencia de los consumidores (Fernández-Vázquez et al., 2011). La industria alimentaria incluye productos de formas y tamaños muy diferentes, y por ello la visión artificial se presenta como una opción óptima para evaluar el color, independientemente de la naturaleza de la muestra. Además del color, se pueden medir características relacionadas con la apariencia pero que afectan a la percepción del ojo humano, como son la textura y la heterogeneidad (Valous et al., 2009). Ya en 1942, David L. MacAdam utilizó elipsoides para marcar regiones en el espacio de color que compartían propiedades comunes (MacAdam, 1942). En ese estudio, MacAdam representó, en términos de desviación estándar, distancias en el espacio de color CIE 1931-(x,y) y ajustó esas regiones a elipsoides. Más tarde, otros autores utilizaron redes compactas de puntos para la división de esos elipsoides en regiones de color diferenciables por el ojo humano (Judd & Wyszecki, 1975; Luke, 1999; Wyszecki, 1954).

Como el color puede extraerse de cada píxel de la imagen, han surgido muchas variables con el fin de cuantificar la heterogeneidad cromática. La mayoría

de los estudios la representan simplemente con la desviación estándar de las variables CIELAB de la región de interés (Mendoza et al., 2006; Valous et al., 2009; Yam & Papadakis, 2004), y esta desviación estándar es en mayor medida producto de la dispersión del color que del error de la medida. Además de la desviación estándar, existen otras magnitudes escalares para su cuantificación, como la media de las diferencias de color con respecto al color medio (MCDM) (Berns, 2000) y la entropía, que es adimensional (Arzate-Vázquez et al., 2011). Otros autores, al intentar explicar más características de la forma que esa dispersión puede tener en el diagrama de color, recurren a representar gráficamente las nubes de puntos, que si bien no pueden cuantificarse, permiten una discusión más profunda que una simple desviación estándar. Además, si la dispersión de los puntos en el espacio de color tiene forma alargada, la desviación estándar de las variables colorimétricas depende de la orientación de la nube. Es decir, nubes de puntos iguales, que tienen la misma heterogeneidad, pero con distintas orientaciones tendrán diferente desviación estándar de sus coordenadas dependiendo de dicha orientación.

Otro desafío al evaluar el color surge cuando el objeto está compuesto por dos o más nubes de puntos diferentes. En estos casos puede ocurrir que el color medio ni siquiera se encuentre realmente presente en la muestra. Actualmente no hay procedimientos para discernir objetivamente el número de grupos de colores independientes que existen en una muestra. Algunos autores han utilizado técnicas de clusterización para segmentación o detección en alimentos (Li et al., 2012), pero no para clasificar nubes de puntos de color en análisis de imagen.

### **2.2. Objetivos**

Se propone, como objetivo general, la mejora de la cuantificación y la comprensión de la heterogeneidad de color en muestras de alimentos mediante la aplicación de elipsoides, dentro del espacio de color CIELAB, para el ajuste de las nubes de puntos de color obtenidas por análisis de imagen. Se ha establecido una metodología fácil de ejecutar desde un punto de vista computacional para la

evaluación de alimentos de color heterogéneo. Los objetivos concretos planteados en este trabajo fueron:

- Uso de técnicas de clusterización para la dilucidación del número de colores presentes en una muestra.
- Desarrollo del algoritmo para el ajuste de las nubes de puntos a elipsoides de color.
- Aplicación de la metodología con fines predictivos.
- Establecimiento de un índice de maduración para la asignación de fases en las semillas de uva durante la maduración a partir de sus características colorimétricas.

### **2.3. Parte experimental**

Para asegurar la obtención de nubes de puntos de diferentes magnitudes fue necesario realizar el estudio sobre muestras que presentaran una superficie con notable heterogeneidad cromática. Así, como ejemplos diversos, se seleccionaron alimentos como manzanas, naranjas, coles y tomates de variedades aparentemente heterogéneas. También se utilizaron uvas de la variedad Zalema y semillas de uvas de las variedades Tempranillo, Syrah y Zalema.

La metodología se aplicó en las uvas con fines de predicción de la madurez tecnológica a partir de las características de los elipsoides de color propuestos. También se aplicó la metodología para proponer un índice de madurez de las semillas de uvas según sus características cromáticas

#### **2.3.1. Construcción de los elipsoides y evaluación de la capacidad predictiva**

Para la puesta a punto de la metodología se utilizaron las muestras de alimentos seleccionados, con heterogeneidad cromática. En todos los casos, las

coordenadas CIELAB se obtuvieron a partir de las imágenes adquiridas con el sistema DigiEye®.

El criterio de segmentación utilizado debe ser, en cada caso, el adecuado a los objetivos del estudio. En este estudio en concreto, la metodología desarrollada debía ser útil en muestras con características cromáticas variadas. Se eligió un criterio de segmentación basado en umbralización diferente al utilizado en el capítulo anterior, que selecciona como región de interés los píxeles de la imagen cuyo cociente  $C_{ab}^*/L^*$  fuera mayor a 2.6 unidades CIELAB. Este umbral excluía el fondo usado en las imágenes, de alta claridad y un croma muy bajo, y a la vez era suficientemente permisivo para incluir en la región de interés todos los tipos de alimentos analizados.

A partir de la nube de dispersión cromática de una muestra, se aplicó un método de clusterización, propuesto por Yager y Filev (1994), para determinar el número de grupos de color independientes. Es un método sustractivo que calcula la probabilidad de que cada punto de la nube sea el centro de un clúster, asumiendo que cualquier punto de la distribución puede serlo. Este método permite modificar el grado de influencia para diferenciar si un conjunto de puntos pertenece a uno o dos grupos. En este caso, se utilizó como límite la diferencia de color mínima distinguible por un ojo humano no entrenado (Martínez et al., 2001).

Cada clúster estaba formado por una matriz ( $m \times 3$ ), que contenía las coordenadas colorimétricas de cada punto. Para calcular la orientación preferente de cada grupo de puntos, se calculó la pendiente de la recta que mejor se ajustaba a los puntos mediante regresión lineal robusta por mínimos cuadrados, la cual elimina los efectos de posibles *outliers*. Tras calcular dicha orientación, para alinear la nube de puntos con los ejes de coordenadas, se multiplicó cada fila de la matriz inicial por una matriz de rotación que corrigiera la dirección calculada mediante la pendiente del ajuste anterior. Esta nueva posición de la nube de puntos carece de sentido colorimétrico, pero a partir de ella es posible calcular las desviaciones estándar de las nuevas coordenadas, que serían útiles para establecer las dimensiones del elipsoide.



A partir de los datos obtenidos, se creó un elipsoide que definiera cada nube de puntos, cuyo centro coincidía con la media aritmética de dicha nube y estaba orientada según la pendiente calculada. Las dimensiones de dicho elipsoide se fijaron en dos veces la desviación estándar de las coordenadas de los puntos tras alinearlos con los ejes. Se consideró el doble de la desviación estándar, que comprende el 95% de los valores en una distribución normal. Se calculó también el número de colores discernibles que encerraba cada elipsoide de color. Para ello, se construyó dentro del elipsoide una red compacta de puntos, que estuvieran separados 3 unidades CIELAB, distancia dentro del espacio que es considerada la diferencia de color mínima que un ojo humano puede percibir (Martínez et al., 2001).

Se estudió cómo afectaba la resolución a la dispersión cromática de las imágenes y al resultado de la clusterización según el color variara drásticamente de unas zonas de la muestra a otras o bien fuera degradado. Se examinó también cómo la metodología propuesta analizaba el color en las semillas de uva durante la maduración. Se vio cómo cuando sucede el pardeamiento en estadios intermedios, el algoritmo diferenciaba las partes claras y oscuras de la semilla y evaluaba su color independientemente.

Como el propósito de la metodología era poder usar las características de los elipsoides con fines predictivos, se utilizó en uvas blancas para la estimación del contenido en azúcares, parámetro habitualmente utilizado en enología como indicador de la madurez. Para ello, se consideró un total de 254 uvas blancas de la variedad Zalema, muestreadas en nueve fechas a lo largo del intervalo de tiempo en el que tiene lugar el envero, y se midieron las características de sus nubes de puntos mediante el método de elipsoides de color propuesto. Como referencia, se midió la concentración de azúcares con por refractometría según el método de la Organización Internacional de la Viña y el Vino (OIV, 2013a). Como técnica quimiométrica para la predicción se utilizó la regresión por mínimos cuadrados parciales (PLSR), donde el contenido de azúcares se usó como variable dependiente (matriz  $Y$ ) y las características de las nubes de puntos se utilizaron como variables dependientes (matriz  $X$ ). Fue posible predecir la concentración de

---

azúcares (que estuvo comprendida entre 5.5 y 18.0 °Brix) con un error (RMSECV) de 0.5 °Brix, y se alcanzó un coeficiente de determinación  $R^2=0.960$  para la validación cruzada.

### 2.3.2. Propuesta de un índice de maduración de semillas basado en sus características colorimétricas

Otro de los objetivos del estudio fue la modelización de un índice de maduración que estableciera fases en las semillas de uva y que utilizara como variables las características de los elipsoides de color para delimitar dichas fases. De un total de 7635 semillas en diferentes estados, se midieron las coordenadas colorimétricas de sus píxeles, tanto en el espacio RGB como en CIELAB. También se consideraron como variables las desviaciones estándar de las coordenadas, el MCDM y todas las variables que caracterizaban el elipsoide de color a partir de las nubes de puntos. Con todas ellas, se creó un espacio multidimensional con treinta y una variables.

Para pre-establecer niveles de maduración que sirvieran como base en el modelo, se recopilieron manualmente 1114 semillas en cuatro estados claramente diferenciables y descritos por Ristic et al. (2005). Estos estados fueron denominados preliminarmente: “verdes”, “amarillas”, “intermedias” y “pardas” (Figura 28). Para evitar el posible error cometido durante la clasificación visual, se eliminaron las semillas que no estuvieran en el intervalo del valor medio más y menos una desviación estándar, al menos para una variable, seleccionándose finalmente 495 semillas de las 1114 iniciales.

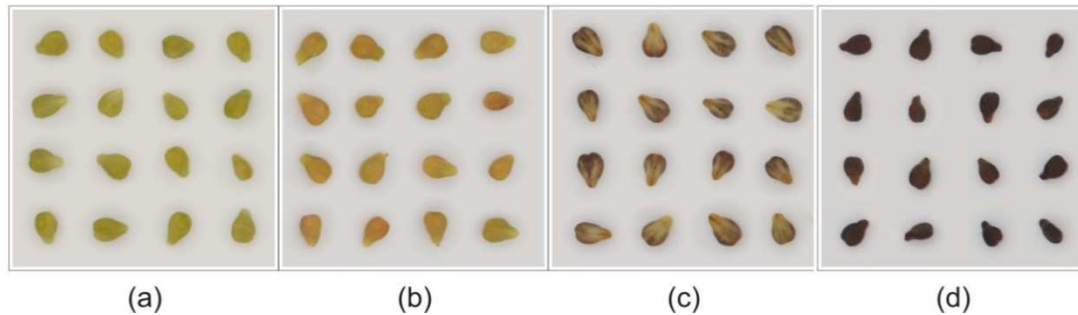


Figura 28. Apariencia de algunas semillas elegidas inicialmente para el modelo. (a) verde, (b) amarillenta, (c) intermedia, (d) parda.

Fijados los valores de cada variable en cada nivel preliminar, se asignó a cada fecha de muestreo el nivel cuya distancia euclídea en el espacio multidimensional fuera mínima.

Un modelo con solo cuatro niveles, de los cuales los dos primeros corresponden a estadios muy tempranos, es claramente insuficiente desde el punto de vista de la práctica enológica. Se hace necesario, por tanto, conseguir una clasificación más fina del nivel de maduración fenólica de las semillas de uva para tomar decisiones robustas y objetivas sobre fechas vendimia según su destino en vinificación.

Por ello, se programaron distintos procedimientos (lineales y no lineales) para la modelización del proceso de maduración a través de una redefinición de los límites de cada nivel, más bien con el fin de probar la robustez de los métodos de clasificación. Así, se consiguieron diferentes alternativas a la escala desde las más sencillas con centroides equidistantes, tanto temporal como colorimétricamente, hasta las más complejas, basadas en modelos no lineales con promedios no equidistantes.

La Figura 29 muestra el resultado de uno de los algoritmos lineales ensayados. Con este modelo se pudieron establecer siete niveles que resultaron de la inter y extrapolación de los cuatro preliminares dentro del espacio multivariante considerado. En este caso se consideraba un intervalo temporal en la toma de muestra de unos diez días, lo que resultó más coherente con los criterios enológicos.

Para comprobar la eficacia del método, una vez establecidos dichos niveles de maduración, se procesaron de nuevo las imágenes asignando a cada semilla la etapa a la cual la distancia euclídea multidimensional fuera mínima. En la figura se muestran ejemplos de secciones de imágenes y la asignación de las semillas a los diferentes niveles de maduración establecidos por el programa (el código de colores utilizado tiene únicamente fines de ilustración para la clasificación de las semillas).

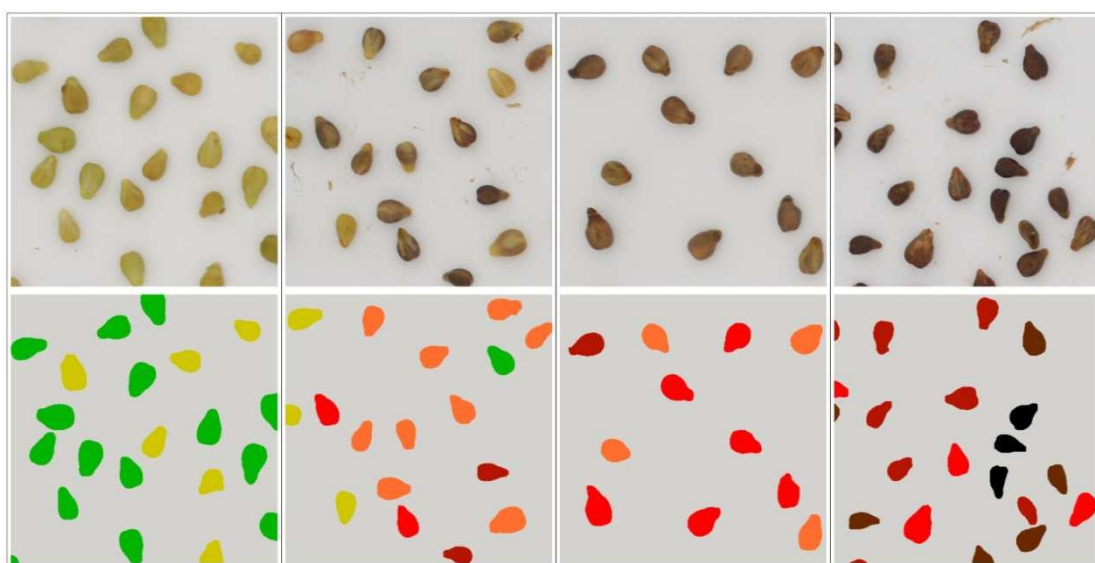


Figura 29. Aplicación de la asignación de semillas por niveles de maduración. verde (I), amarillo (II), naranja (III), rojo (IV), rojo oscuro (V), marrón (VI), negro (VII).

La Figura 30 muestra la evolución de algunos parámetros colorimétricos en los distintos niveles de maduración. El resultado más obvio fue el descenso de las variables con significado más cuantitativo (claridad y croma) provocado por el pardeamiento progresivo de las semillas. Atendiendo al tono, como componente del color de índole cualitativa, se observó cómo al comienzo de la maduración, las semillas se encontraban en el segundo cuadrante del plano  $a^*b^*$ , y fueron cambiando desde tonos verdosos hacia más amarillentos hasta alcanzar valores de  $40^\circ$ , pero donde apenas era apreciable visualmente el tono debido a los bajos valores de croma y de claridad. Como describe la bibliografía, y se pudo apreciar en la evaluación visual, el pardeamiento no sucede homogéneamente, sino que aparece una zona oscura en la parte más ancha de la semilla, y se extiende hasta

pardearla por completo. Esta heterogeneidad se tradujo en un aumento de la desviación estándar de la claridad en los niveles intermedios.

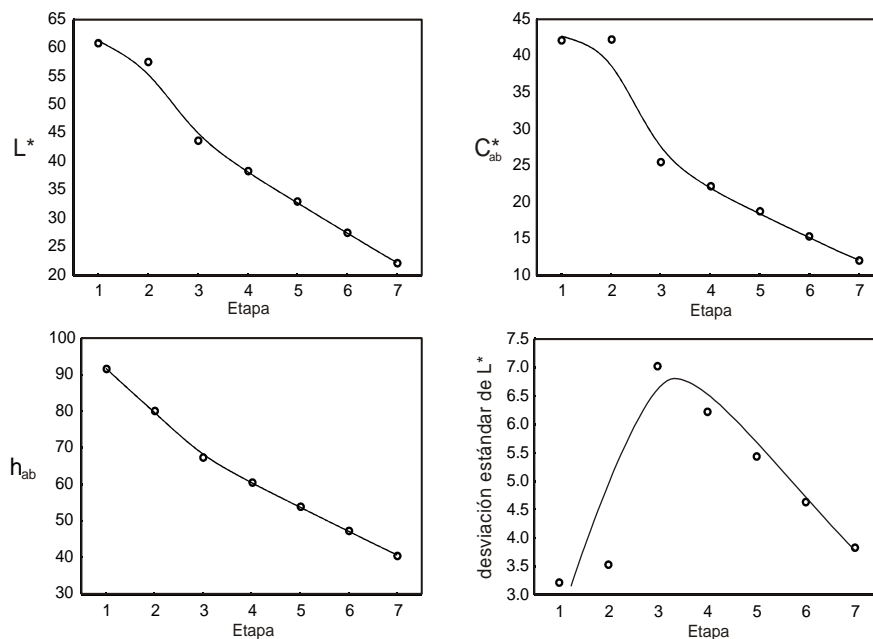


Figura 30. Evolución de las características colorimétricas a lo largo del índice de maduración propuesto.

Como era esperable, en cada muestreo se encontraron semillas con diferentes niveles de maduración. La presencia relativa de cada una de ellas ocurrió según la variedad y la fecha de muestreo (Figura 31). Para la variedad Zalema, los niveles 1 y 2 predominaron en los estadios iniciales y, a partir de finales de julio, las semillas comenzaron a pardearse de manera heterogénea; de hecho, las muestras de las dos últimas semanas de agosto contenían semillas pertenecientes a todos los niveles del modelo. Aplicando el método desarrollado para el establecimiento del nivel de maduración, se encontró que en el momento de la vendimia, dos tercios de las semillas estaban en los estados 5 y 6, y sólo un 10% de las semillas había alcanzado la etapa 7 de maduración.

Los perfiles de las variedades tintas fueron muy similares entre sí, y diferentes al de la variedad blanca. En las tintas, los dos niveles iniciales se encontraron más solapados que en las uvas blancas, y a mitad de verano, ya no se encuentran semillas en ninguno de estos niveles presentando una evolución más homogénea. De hecho, al desaparecer las semillas de los niveles iniciales, el nivel 3

(principio del pardeamiento) fue el estado dominante para Syrah y Tempranillo, a diferencia de Zalema donde sus semillas coexistieron en diferentes niveles al mismo tiempo. Probablemente debido a la temprana vendimia de las variedades tintas, muy pocas semillas habían alcanzado los niveles 6 y 7. Merece la pena destacar que, a pesar de la similitud de los perfiles de distribución de las dos variedades tintas, existió un desfase en el tiempo entre ambas, concretamente de una semana, y coincidió a su vez con el desfase que hubo entre ambas vendimias.

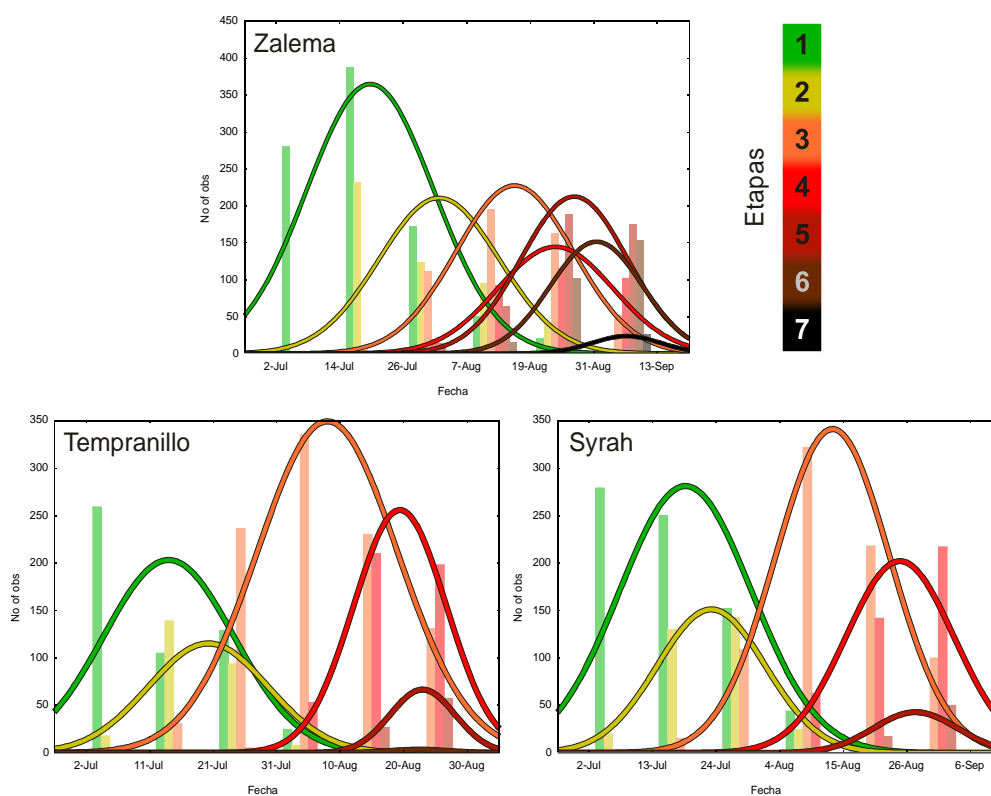


Figura 31. Distribución de las semillas según su nivel de maduración y variedad.

## 2.4. Conclusiones

- Se ha probado un algoritmo para el cálculo del número de grupos de color, y por tanto regiones con diferente composición, que están presentes en una muestra de alimentos. Este método sustractivo de clusterización se basa en el umbral de percepción del ojo humano.
- El método ajusta cada nube de puntos a un elipsoide cuya posición, orientación y dimensiones explica la forma de dicha nube, y ha sido aplicado

con éxito a muestras de alimentos con diferentes atributos de apariencia: forma, tamaño, color y textura.

- Las características de los elipsoides calculados explican el concepto de heterogeneidad cromática mejor de lo que lo hace el uso de una simple desviación estándar de las coordenadas colorimétricas.
- Se ha probado con éxito la estimación de parámetros químicos a partir de las características de los elipsoides calculados en muestras, mediante el uso de técnicas estadísticas multivariantes.

La comprensión de la heterogeneidad cromática, tal como se evalúa en este trabajo, ha permitido la propuesta de un índice de maduración de semillas que tiene en cuenta la forma de sus nubes de puntos. Al asignar este índice de maduración a una cantidad considerable de semillas, se han observado patrones de evolución similares en las variedades tintas Tempranillo y Syrah, y diferentes al de la variedad Zalema, lo que sugiere que la apariencia que tienen las semillas de una muestra no es trivial y está relacionada con su evolución natural.







## Analysis of food appearance properties by computer vision applying ellipsoids to colour data



Francisco J. Rodríguez-Pulido, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia\*

Food Colour & Quality Lab., Dept. Nutrition & Food Science, Facultad de Farmacia, Universidad de Sevilla, 41012 Sevilla, Spain

### ARTICLE INFO

#### Article history:

Received 4 July 2013

Received in revised form 27 August 2013

Accepted 31 August 2013

#### Keywords:

Food colour  
Colour ellipsoids  
CIELAB  
Image analysis  
Clustering

### ABSTRACT

The use of computer vision for estimating composition in food products has become wide spread in recent years, especially for products where by measuring colour or other spectral features, we are able to estimate the composition, variety, or ripeness. On the other hand, the appearance is one of the most worrying issues for producers due to its influence on quality and consumer preferences. Computer vision is the best option to satisfy the need of measuring colour and heterogeneity in these products. Previous studies have expressed the heterogeneity with the standard deviation or other magnitudes that do not explain accurately the distribution of colour in CIELAB colour space. Graphing the scatterplot of the  $a^*b^*$  values belonging to the pixels of the image helps to explain the shape of the point cloud, but currently there is not an objective procedure to quantify these point clouds. This work has established a method for improving the illustration of colour measurements by image analysis. The proposed algorithm classified the point clouds by clustering methods and established the methodology for fitting the resulting clusters into ellipsoids. Their geometric features described the shape of the clouds in a quantitatively manner and they could be useful in multivariate statistical techniques for classifying and predicting chemical properties.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

Computer vision has radically changed the outlook of evaluating the composition of food products in recent years. Its advances have improved within the framework of accuracy, and joined to chemometrics, 'Chemical Imaging' is widely used nowadays (ElMasry and Sun, 2010). The higher requirements on quality control has entailed that it is not enough analysing only the chemical composition but also the spatial distribution within a sample (Du and Sun, 2004). The relationship between chemical composition and spectral properties have been well studied in infrared and visible spectra by several techniques, such as near infrared spectroscopy, near infrared reflectance spectroscopy, and visible and infrared hyperspectral imaging (Hernández-Hierro et al., 2012; Rodríguez-Pulido et al., 2012; Barbin et al., 2013; Mathiassen et al., 2011; Romano et al., 2012). Besides of composition, these measurements are important since food appearance is one of the most important characteristics due to its relationship with quality and consumer preferences (Fernández-Vázquez et al., 2011; Calvo et al., 2001). Since food industry includes products having very different sizes and shapes, computer vision arises as a suitable option

to satisfy the need of measuring colour regardless the nature of samples. Moreover, computer vision allows measuring not only colour but also other features related to appearance, that do not vary the colour, but they affect to how the human eye perceive it, such as texture or heterogeneity (Valous et al., 2009).

Back in 1942, David L. MacAdam used ellipsoids for marking regions in colour spaces having common properties (MacAdam, 1942). In that study, the standard deviations were represented in terms of distance in the CIE 1931 colour space chromaticity diagram and these regions were fitted into ellipses. They showed that sources radiating spectral distributions belonging to these regions of the diagram appeared to have the same colour, for the average human observer. Later, some authors used a "closest packing" lattice of points to improve the understanding of the space involved by ellipsoids (Salmerón et al., 2012; Wyszecki, 1954; MacAdam, 1974; Luke, 1999; Judd and Wyszecki, 1975). In these cases, the lattice constant might represent the smallest number of just-noticeable chromaticity steps between the two chromaticities represented by the two points.

In computer vision, device-dependent colour spaces are commonly employed in image analysis because this kind of information is given by cameras directly (Jack, 2008). Nevertheless, colour must be measured by device-independent colour spaces (its appearance does not depend on the device) to ensure the

\* Corresponding author. Tel.: +34 954556495; fax: +34 954556110.  
E-mail address: [heredia@us.es](mailto:heredia@us.es) (F.J. Heredia).

objectivity of the measure. Among these spaces, CIELAB is one of the colour spaces recommended by the International Commission on Illumination (CIE) and it is considered perceptually uniform, meaning that just-detectable visual difference constitutes a constant distance in any location or direction within the space (CIE, 1976). Therefore CIELAB is widely used as a standard space for comparing colours because of its reliability.

Within CIELAB, a psychometric index of lightness ( $L^*$ ) and two colour coordinates ( $a^*$  and  $b^*$ ) are defined.  $L^*$  is the quantitative attribute of relative luminosity, which is the property according to which each colour can be considered as equivalent to a member of the grey scale, ranging between black ( $L^* = 0$ ) and white ( $L^* = 100$ ). Coordinate  $a^*$  takes positive values for reddish colours and negative values for greenish ones, and  $b^*$  takes positive values for yellowish colours and negative values for bluish ones. From the Cartesian coordinates ( $a^*$  and  $b^*$ ), two angular parameters can be defined: chroma and hue or hue angle. Hue angle ( $h_{ab}$ ) is the qualitative attribute that allows any colour to be graded as reddish, greenish, etc., and chroma ( $C_{ab}^*$ ) is considered the quantitative attribute of colourfulness, allowing assessing the degree of difference of any given hue relative to a grey colour with the same lightness (Hutchings, 1999).

Obtaining the CIELAB coordinates by image analysis requires a camera which records visible light in gradations of three basic colours: red, green and blue (RGB). This device-dependent colour space may be transformed into CIELAB coordinates by means of a calibration which in turn requires controlled lighting (CIE, 2007; León et al., 2006). After taking images and transforming between colour spaces, a segmentation criterion is applied for calculating colour only from pixels with analytical information, also known as region of interest (ROI). There are different techniques of segmentation, being thresholding and edge-detection the main ones (Cheng et al., 2001; Zheng and Sun, 2008).

Since colour can be extracted from each pixel of the ROI, some variables emerge in order to express the degree of heterogeneity. Most of the studies show the result of measurements as the average colour and its standard deviation from all the pixels selected of the ROI (Yam and Papadakis, 2004; Valous et al., 2009; Mendoza et al., 2006; Girolami et al., 2013; Dufossé et al., 2005), being this standard deviation mainly a consequence of the heterogeneity instead of the measuring error. Besides the standard deviation, there are more scalar magnitudes in order to quantify the heterogeneity of samples, such as the mean colour difference from the mean (MCDM) (Berns, 2000) and entropy, which is dimensionless (Arzate-Vázquez et al., 2011). Further than a simple point plus a scalar explaining its heterogeneity, some authors resort to graph the scatterplot of the  $a^*b^*$  values as a point cloud, corresponding the points to the colour of each pixel (Urban et al., 2007; Palus, 2006). This option is quite useful, because these graphs are fairly explanatory and give an idea about the colours present in the sample as well as the relative presence of each one. However, a problem arises when the colour of a sample is spread out in two or more different point clouds. In these cases, the average colour may represent a point that cannot even be present in the sample. Currently, there are not objective procedures for discerning how many groups of colours are present in a sample. Some authors have used clustering methods on image analysis, not for classifying colours but for segmentation or detection purposes (Li et al., 2012; Yin et al., 2011).

This work has established an easy to carry out methodology for improving the evaluation of heterogeneous colours in food products and the illustration of these measurements in CIELAB colour space. The proposed algorithm could be useful for obtaining analytical information in studies where by chemometrics, the relationship among colour, appearance, and composition wants to be studied.

## 2. Materials and methods

### 2.1. Imaging system

For acquiring images, the DigiEye<sup>®</sup> system based upon a calibrated digital camera was used (Luo et al., 2001). It includes an illumination box specially designed by VeriVide Ltd. (Leicester, UK) to illuminate the samples consistently and a digital camera connected to a computer. The digital camera used for image acquisition was a 10.2-megapixel Nikon<sup>®</sup> D80 with Nikkor<sup>®</sup> 35 mm f/2D objective. It was calibrated with the DigiTizer (VeriVide Ltd., Leicester, UK) colour chart with the aim of characterise the camera response by relating its RGB signals to CIE specifications. The cabinet is equipped with two fluorescent tubes that emulate the CIE standard illuminant D65 and offer stable lighting conditions (CIE, 2007). They were switched on at least 10 min before being used, according to manufacturer indications, to stabilize them.

The application of the methodology and the algorithm for computing the ellipsoids from point clouds were developed on MATLAB (The Mathworks, 2009). Within MATLAB, the Fuzzy Logic, Image Processing, Partial Differential Equation and Statistics Toolboxes were also used.

The aim of this work was the establishment of a new methodology for the interpretation of the colour heterogeneity, so the development of the method and its application to different food samples have been included in Section 3.

### 2.2. Samples

The algorithm was applied to food products having different size and appearance for showing the results in a clear manner. Cabbages (*Brassica oleracea* var. *Viridis*), oranges (*Citrus sinensis* L. Osbeck var. *Navelate*), apples (*Malus domestica* var. *Kanzi*), and tomatoes (*Solanum lycopersicum* L. var. *Kumato*) purchased from local retailers were studied. Seeds from red grapes (*Vitis vinifera* var. *Syrah*) were included as representative of small-size samples. These fruits and vegetables were chosen based on the representativeness of foodstuff having different colour and heterogeneity. Homogeneous food products such as wines or juices were not considered since the evaluation of heterogeneity in this type of samples has not sense.

For proving the potential of the method, the sugar content of grapes, which is an indicator of maturity in oenology, was estimated by means of the proposed method. For this purpose, 254 white grapes (*V. vinifera* var. *Zalema*) were taken at nine dates during the interval of time where the change of colour occurs. The vineyards sampled are included under the “Condado de Huelva” Designation of Origin, in Southwestern Spain, harvested in 2012. As reference method, an Abbe refractometer was used to evaluate the sugar concentration according to the method of The International Organisation of Vine and Wine (OIV, 2013).

## 3. Results

### 3.1. Methodological procedure

#### 3.1.1. Image acquisition

As is usual in computer vision, the images were acquired under diffuse illumination for avoiding undesired glints and shadows. The background were chosen for make easier the segmentation process, to the extent possible. A thick sheet was used for this purpose since it is considered a good Lambertian surface (diffuse reflectance surface which does not vary depending on the viewing angle) (Jaglarlz et al., 2006).

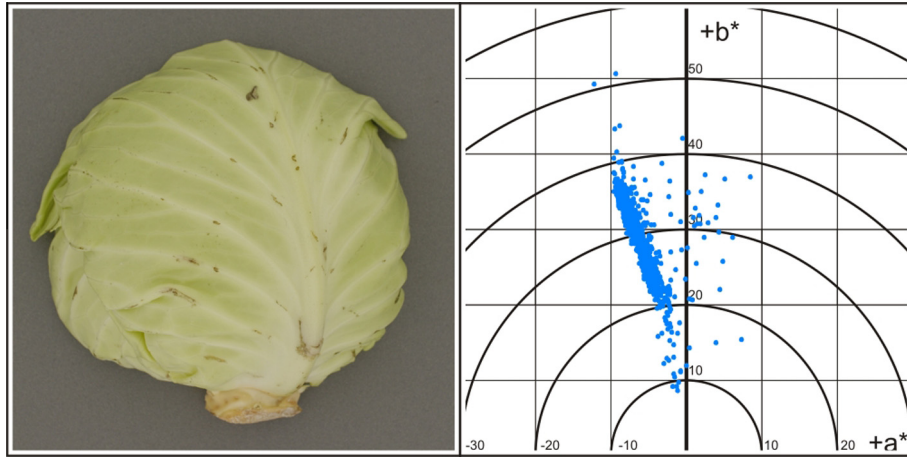


Fig. 1. Image of a cabbage and its colour on the  $(a^*b^*)$ -plane.

3.1.2. Segmentation

The segmentation criteria depend on nature of samples. In this case, and based on the chromatic differences among samples and background, the pixels having the ratio  $C_{ab}^*/L^*$  higher than 2.6 CIE-LAB units was chosen as segmentation criterion. Then, the inner holes in segmentation mask were filled, and the resulting regions were eroded in order to avoid possible aberration of colours at the edge of the objects.

3.1.3. Clustering

The methodology proposed by Yager and Filev (1994) was followed for the clustering process. This subtractive clustering method assumes that each point is a potential cluster centre and calculates a measure of likelihood that each data point would define the cluster centre, based on the density of surrounding data points. It can be summarised in three steps. In the first step it discretizes the object space and in doing so, it generates the potential

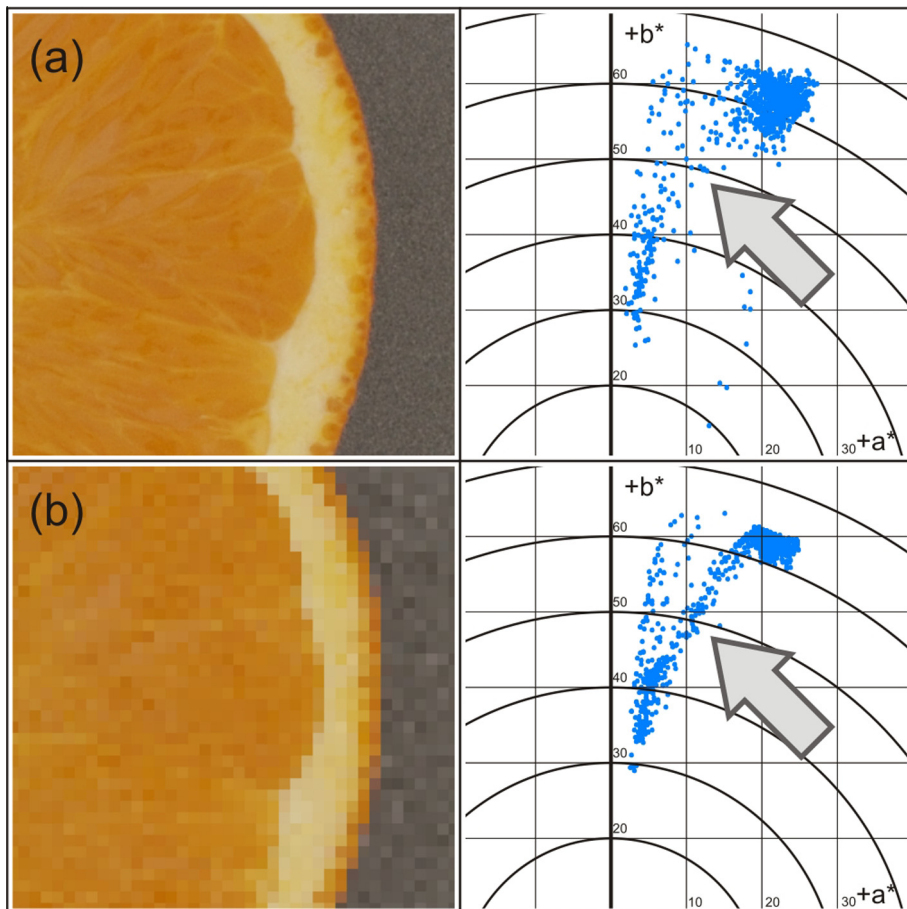


Fig. 2. Point clouds of a slice of orange depending on the resolution of image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

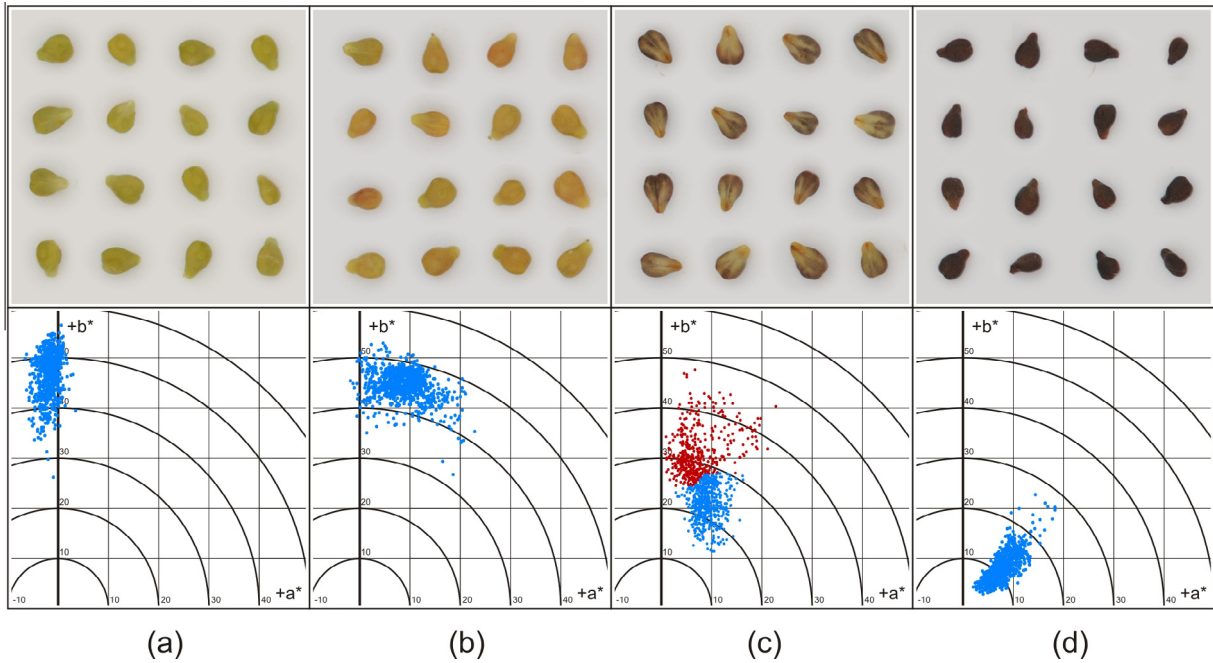


Fig. 3. Image of grape seeds at different stages of ripeness and their point clouds after applying the clustering process.

cluster centres. For locating the next data cluster and its centre position, the second step removes all data points in the vicinity of the first cluster centre. Finally, it iterates on this process until all data are under the influence of a cluster centre. Due to all data were included in the same bidimensional space, the range of influence was set in agreement with the acceptable tolerance by the human eye (Martínez et al., 2001).

3.1.4. Fitting to ellipsoids

Each cluster was composed by a  $(m \times 3)$ -matrix containing the CIELAB colorimetric variables of the points:

$$\text{cluster}_n = \begin{pmatrix} L_1^* & a_1^* & b_1^* \\ L_2^* & a_2^* & b_2^* \\ \vdots & \vdots & \vdots \\ L_i^* & a_i^* & b_i^* \\ \vdots & \vdots & \vdots \\ L_m^* & a_m^* & b_m^* \end{pmatrix} \quad (1)$$

From each cluster, the orientation of the ellipsoid with respect to the  $(a^*b^*)$ -plane was calculated by robust linear regression of the points. This regression improved the result calculated by simple linear regression, removing the effects of outliers. The orientation, in degrees, was calculated by the expression:

$$\theta = \text{atan } b \quad (2)$$

being  $b$  the slope obtained by the regression. Each row of the matrix was multiplied by the rotation matrix  $R$ . The new points  $(L^*, a'', b'')$  had not colorimetric sense, but they were useful for calculating the axes of the ellipsoid:

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \quad (3)$$

$$\begin{pmatrix} L_i^* & a_i^{*'} & b_i^{*'} \end{pmatrix} = \begin{pmatrix} L_i^* & a_i^* & b_i^* \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \quad (4)$$

Thus, the centre of the ellipsoid was located at:

$$\text{Centre} = (\bar{L}^*, \bar{a}^*, \bar{b}^*) \quad (5)$$

and the dimensions of the ellipsoid along the direction defined by  $\theta$  were defined by:

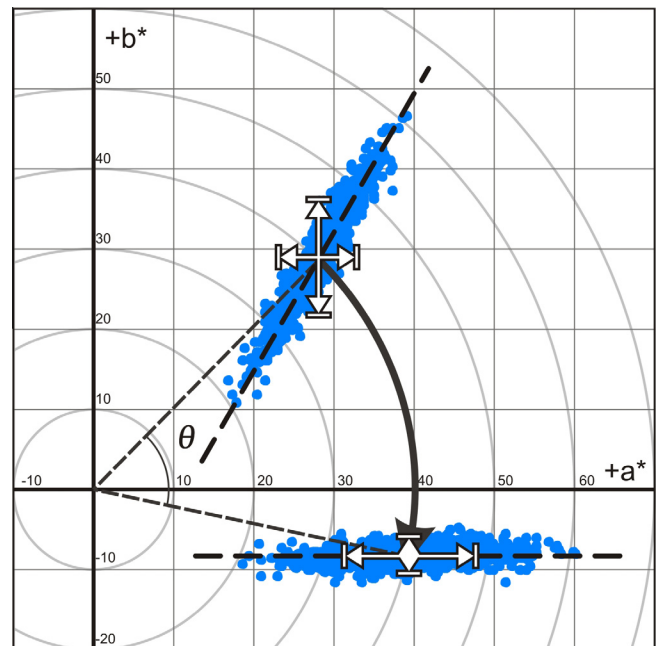


Fig. 4. Point cloud showing the rotation operation within the  $(a^*b^*)$ -plane. White arrows show the standard deviation along two axes before and after the rotation.

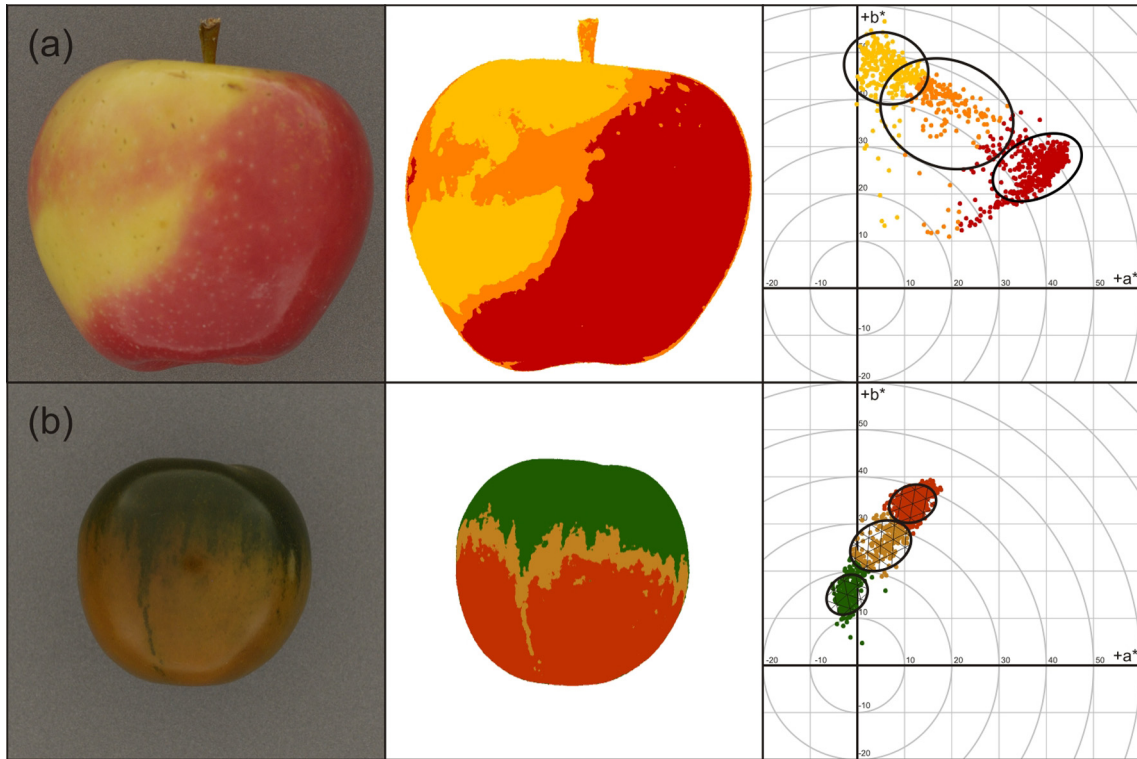


Fig. 5. Images of food products after applying the clustering and fitting processes.

$$\text{Length} = 2 \times \left[ \frac{1}{m-1} \sum_{i=1}^m (a_i^{*f} - \bar{a}^{*f})^2 \right]^{\frac{1}{2}} \quad (6)$$

$$\text{Width} = 2 \times \left[ \frac{1}{m-1} \sum_{i=1}^m (b_i^{*f} - \bar{b}^{*f})^2 \right]^{\frac{1}{2}} \quad (7)$$

$$\text{Height} = 2 \times \left[ \frac{1}{m-1} \sum_{i=1}^m (L_i^* - \bar{L}^*)^2 \right]^{\frac{1}{2}} \quad (8)$$

The axes of the ellipsoid were set to two standard deviations, where about 95% of the values lie within this scope in a normal distribution.

### 3.2. Application of the methodology to food products

The method was applied to the following products:

For an image of a cabbage, the point cloud on the ( $a^*b^*$ )-plane of the points is shown in Fig. 1. The colour of cabbage varied from greenish (with positive values for  $b^*$  and negative ones for  $a^*$ ) to achromatic ones (values close to zero for  $a^*$  and  $b^*$ ). In this case, there was a colour gradation between both areas.

Nevertheless, colour might change drastically among different areas in images. In this respect, the role of the resolution of images has to be highlighted. Images having high resolution will have fewer points between point clouds than images having low one. Fig. 2 shows the point cloud of an orange slice. Almost all the points are concentrated in the region of the diagram that represents typically orange colours. The white mesocarp, more achromatic, appears closer to the origin of coordinates in a small cloud. In this case, the possibility that a pixel belongs to the boundary between the endocarp and mesocarp is higher in the low resolution image. Additionally, the extension of the points belonging to orange area reduces its size in low resolution images because the representativeness of colour decreases as well (Fig. 2b).

Another example including the clustering process is shown in Fig. 3. Images of grape seeds at different stages of ripeness and their point clouds after applying the clustering process are shown in Fig. 3. Fig. 3a, b, and d were composed of one cluster each one. Conversely, Fig. 3c showed more heterogeneity than the others. This way, if it had been considered as a unique point cloud, it would have had a higher standard deviation, but its belonging to two clusters would not have been explained. Because this sample was composed of two independent colours, it must be considered the individual heterogeneity of each one.

Table 1  
Cluster centres and dimensions of the fitted ellipses for images of Fig. 5. ( $a^*$ ,  $b^*$ , and size of ellipse in CIELAB units).

	Cluster	Area on image (%)	$a^*$	$b^*$	Ellipse length	Ellipse width	Eccentricity	Orientation angle (°)
Apple	Red	53.40	38.1	25.6	10.1	6.3	0.78	26.9
	Orange	17.90	19.0	37.0	14.4	11.3	0.56	-20.9
	Yellow	28.70	6.1	46.6	9.0	7.5	0.62	-17.0
Tomato	Red	45.60	11.7	34.3	5.1	3.9	0.65	20.4
	Pale brown	19.37	4.9	25.4	6.7	5.0	0.67	26.3
	Green	35.03	-2.2	15.0	4.7	3.8	0.59	41.6

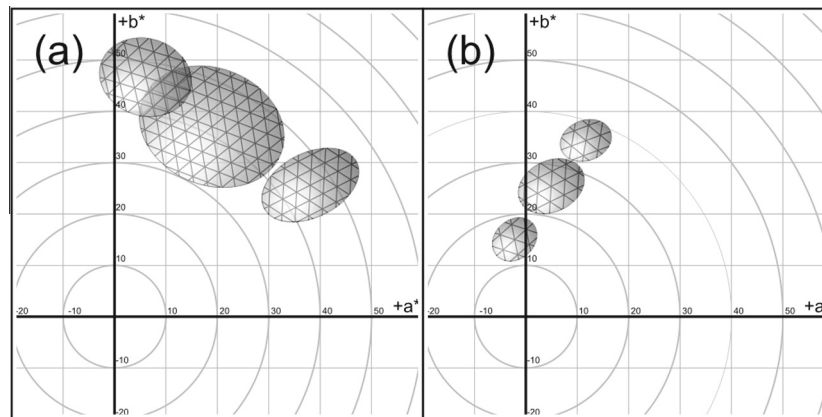


Fig. 6. Orthographic projection of ellipsoids and spatial lattice on the plane  $a^*b^*$ . The point clouds have been removed.

Table 2

Centres and dimensions for ellipsoids of Fig. 6. ( $L^*$ ,  $a^*$ ,  $b^*$ , length, width, and height in CIELAB units. Volume in CIELAB units<sup>3</sup>).

Cluster	$L^*$	$a^*$	$b^*$	Length	Width	Height	Volume	Points in lattice	
Apple	Red	37.8	38.1	25.6	10.1	6.3	11.4	379.1	59
	Orange	64.2	19.0	37.0	14.4	11.3	10.2	871.3	136
	Yellow	53.9	6.1	46.6	9.0	7.5	16.8	594.8	93
Tomato	Red	34.0	11.7	34.3	5.1	3.9	7.4	77.2	12
	Pale brown	20.1	4.9	25.4	6.7	5.0	8.0	74.9	11
	Green	26.9	-2.2	15.0	4.7	3.8	6.8	119.1	18

Table 3

Centres and dimensions for ellipsoids of Fig. 7. ( $L^*$ ,  $a^*$ ,  $b^*$ , length, width, and height: CIELAB units; orientation angle: degrees; volume: CIELAB units<sup>3</sup>; the remaining magnitudes are dimensionless).

Sampling	Date	°Brix	$L^*$	$a^*$	$b^*$	Length	Width	Height	Eccentricity	Orientation angle	Volume	Points in lattice
1	2-July	5.5	55.6	-10.6	48.1	7.8	4.1	6.5	0.85	-11.7	108.2	17
2	9-July	6.1	57.6	-10.6	48.4	7.1	4.6	6.6	0.76	-20.9	114.5	18
3	22-July	11.7	61.5	-5.4	48.7	8.7	4.5	5.6	0.86	11.2	113.6	18
4	25-July	12.7	62.1	-5.1	46.0	9.4	5.8	7.4	0.79	-20.0	212.2	33
5	29-July	13.6	62.1	-2.8	46.5	9.7	5.7	6.1	0.81	9.9	175.6	28
6	31-July	14.6	63.2	-2.0	46.2	9.6	5.9	5.8	0.79	0.2	173.6	27
7	5-August	16.1	61.3	-0.9	46.1	9.1	6.7	6.4	0.67	-3.9	202.7	32
8	8-August	16.2	62.7	-0.1	45.7	8.4	5.6	6.6	0.75	-3.0	160.7	25
9	12-August	18.0	61.7	0.5	46.9	7.5	5.3	6.8	0.71	-2.7	140.9	22

On the other hand, in both orange and cabbage cases, the scatter of point has an oblique distribution. This fact involves high standard deviation in  $a^*$  and  $b^*$  coordinates. Nonetheless, the points representing the colour in cabbage are apparently aligned. This means that points having the same distribution will have different standard deviations depending on the orientation of the cloud. Therefore standard deviations of colorimetric coordinates are not suitable values to express the heterogeneity of colour because they do not define the actual shape of these point cloud.

Fig. 4 shows how the rotation operation aligned the point cloud with the axes. As previously described, the new position of the cloud had not colorimetric sense, but the standard deviations obtained were more consistent with the true shape of the cloud and they still had CIELAB units since only the orientation of coordinate system was modified.

Fig. 5 shows images of an apple and a tomato concluding the methodology. The algorithm categorised by clusters the point clouds and included the fitted ellipses. The image of the apple was composed by two principal clusters well defined belonging to red<sup>1</sup> and yellow areas. A small region between them showed

how colours gradually went from red to yellow. Moreover, the bigger size of this intermediate ellipse indicated the greater heterogeneity of this cluster. The algorithm also classified the tomato into three clusters. Nevertheless, there was more homogeneity regarding the size and orientation of its ellipses. The information of each ellipse is described in detail in Table 1. It is important to highlight that orientation angle has not relationship with the qualitative attribute of colour. It is used for giving an idea of the shape of the point cloud. For each cluster, the area of image belonging to this cluster, the centre of the ellipse, its dimensions, eccentricity and orientation were defined.

At this point, only variables  $a^*$  and  $b^*$  were taken into account. Thus the proposed ellipses were not still able to explain how large was the point cloud along the third colorimetric variable, lightness ( $L^*$ ). With the aim of evaluating the distinguishable colours that contains each cluster, the ellipses became into ellipsoids by considering the variable  $L^*$ , building this dimension with the same criterion which the others axes were built. The CIELAB colour space is continuous, meaning that there is an infinite number of colours points within every portion of space considered. Nonetheless, there must be a minimal difference of colour ( $\Delta E$ ) for being noticeable by the human eye. According to Martínez et al. (2001), a value around 3.0 CIELAB units should be considered a preliminary estimate of

<sup>1</sup> For interpretation of colour in Fig. 5, the reader is referred to the web version of this article.

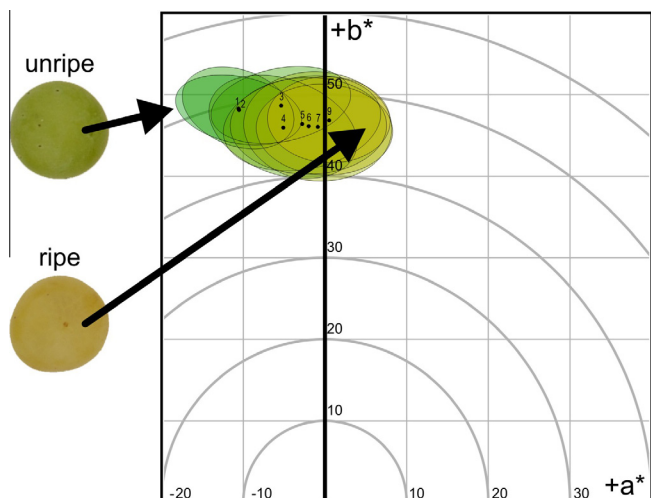


Fig. 7. Evolution of ellipsoids of colour of white grapes during the ripening.  $L^*$  axis has been removed to get a clearer representation.

the acceptable tolerance by a non-trained person in wines. This way, the volume of ellipsoids was filled with a regular rhombohedral (cubo-octahedron) lattice of points. This lattice is a type of “closest packing”, where each point is surrounded by 12 equidistant nearest neighbours, being the distance between points called the *lattice constant*. In our case, this parameter was set at three CIE-LAB units. Fig. 6 shows the ellipsoids obtained for the apple (a) and the tomato (b). The projection of the lattice on the plane  $a^*b^*$  shows the dimension of the lattice constant, where every point represents an individual colour within each cluster. It is worthy remarkable that although the ellipsoids belonging to clusters orange and yellow look overlapped, they actually do not coincide because they are at different heights (different  $L^*$  values). The lower standard deviation of points of clusters in tomato expresses itself as fewer points in the lattices built (Table 2). Anyhow, the lattice constant set at three units could be changed depending on the purpose of the study or the kind of samples in further works.

For testing the ability of the method, the ellipsoids of colour were obtained of grapes as well as their sugar concentration. Table 3 summarises these data regarding the sampling date. The position as well as the shape of ellipsoids changes along the ripening of white grapes, and it is shown in Fig. 7. Then, the sugar concentration was estimated by partial least squares regression (PLSR). Data of sugar content was used as dependent (Y) variable and data of ellipsoids of colour was used as the independent (X) variables in the PLSR. Sugar concentration, which ranged between 5.5 and 18.0 °Brix, had  $R^2 = 0.986$  for calibration and  $R^2 = 0.960$  for cross-validation. In turn, the root mean square error (RMSE) was 0.5 °Brix for calibration and 0.9 °Brix for cross-validation.

#### 4. Conclusions

A novel method for estimating the number of colour groups, and therefore regions having different chemical composition, that are present in an image has been established. This method was based on a subtractive clustering method considering the threshold of perception of the human eye. On the other hand, an objective way for quantifying the resulting point clouds based on the construction of ellipsoids was developed. The orientation and semi-axes of these ellipsoids were faithfully in agreement with the actual shape of the cloud, and improved the explanation of the heterogeneity that a single point and the standard deviation provide. The method was successfully applied in images of foodstuff having different sizes, colours, and textures. Since

heterogeneity may be explained quantitatively by means of new variables, these could be taken into account for being included as new variables in multivariate statistical techniques for classifying and predicting properties in food products.

#### Acknowledgements

This work was supported by the Projects P10-AGR6331 (Consejería de Economía, Innovación, Ciencia y Empresa, Junta de Andalucía), AGL2011-30254-C02 (Ministerio de Economía y Competitividad, Gobierno de España), and the concession of the fellowship BES-2009-025429.

#### References

- Arzate-Vázquez, I., Chanona-Pérez, J.J., de Perea-Flores, M.J., Calderón-Domínguez, G., Moreno-Armendérez, M.A., Calvo, H., et al., 2011. Image processing applied to classification of avocado variety hass (*Persea americana* Mill.) during the ripening process. *Food and Bioprocess Technology* 4, 1307–1313.
- Barbin, D.F., ElMasry, G., Sun, D.W., Allen, P., 2013. Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging. *Food Chemistry* 138, 1162–1171.
- Berns, R., 2000. *Billmeyer and Saltzman's Principles of Color Technology*. Wiley, New York.
- Calvo, C., Salvador, A., Fiszman, S.M., 2001. Influence of colour intensity on the perception of colour and sweetness in various fruit-flavoured yoghurts. *European Food Research Technology* 213, 99–103.
- Cheng, H.D., Jiang, X.H., Sun, Y., Wang, J., 2001. Color image segmentation: advances and prospects. *Pattern Recognition* 34, 2259–2281.
- CIE, 1976. Recommendations on uniform color spaces, color-difference equations, psychometric color terms. Bureau Central de la CIE, Vienna.
- CIE, 2007. Commission internationale de l'Eclairage. Standard Illuminants for Colorimetry. ISO 11664-2, 2007.
- Du, C.J., Sun, D.W., 2004. Recent developments in the applications of image processing techniques for food quality evaluation. *Trends in Food Science and Technology* 15, 230–249.
- Dufossé, L., Galaup, P., Carlet, E., Flamin, C., Valla, A., 2005. Spectrocolorimetry in the CIE  $L^*a^*b^*$  color space as useful tool for monitoring the ripening process and the quality of PDO red-smear soft cheeses. *Food Research International* 38, 919–924.
- ElMasry, G., Sun, D.W., 2010. Meat quality assessment using a hyperspectral imaging system. In: Sun, P.D.-W. (Ed.), *Hyperspectral Imaging for Food Quality Analysis and Control*. Academic Press, San Diego, pp. 175–240.
- Fernández-Vázquez, R., Stinco, C., Meléndez-Martínez, A.J., Heredia, F.J., Vicario, I.M., 2011. Visual and instrumental evaluation of orange juice color: a consumers' preference study. *Journal of Sensory Studies* 26, 436–444.
- Girolami, A., Napolitano, F., Faraone, D., Braghieri, A., 2013. Measurement of meat color using a computer vision system. *Meat Science* 93, 111–118.
- Hernández-Hierro, J.M., Valverde, J., Villacreses, S., Reilly, K., Gaffney, M., González-Miret, M.L., et al., 2012. Feasibility study on the use of visible-near-infrared spectroscopy for the screening of individual and total glucosinolate contents in broccoli. *Journal of Agricultural and Food Chemistry* 60, 7352–7358.
- Hutchings, J., 1999. *Food Color and Appearance*, second ed. Springer.
- Jack, K., 2008. Color Spaces. In: *Digital Video and DSP*. Newnes, Burlington, pp. 15–29.
- Jaglariz, J., Duraj, R., Szopa, P., Cisowski, J., Czernastek, H., 2006. Investigation of white standards by means of bidirectional reflection distribution function and integrating sphere methods. *Optica Applicata* 36, 97–103.
- Judd, D.B., Wyszecki, G., 1975. *Color in Business, Science, and Industry*. Wiley.
- León, K., Mery, D., Pedreschi, F., León, J., 2006. Color measurement in  $L^*a^*b^*$  units from RGB digital images. *Food Research International* 39, 1084–1091.
- Li, C., Wang, B., Wang, J., Li, F., 2012. Extracting vein of leaf image based on K-means clustering. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering* 28, 157–162.
- Luke, J.T., 1999. OSA Instrumental in development of the uniform color scales. *Optics and Photonics News* 10, 28–33.
- Luo, M.R., Cui, G.H., Li, C., 2001. British Patent Entitled Apparatus and Method for Measuring Colour (DigiEye System). Derby University Enterprises Limited (4-10-2001).
- MacAdam, D.L., 1942. Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America* 32, 247–273.
- MacAdam, D.L., 1974. Uniform color scales. *Journal of the Optical Society of America* 64, 1691–1702.
- Martínez, J.A., Melgosa, M., Pérez, M.M., Hita, E., Negueruela, A.I., 2001. Note. Visual and instrumental color evaluation in red wines. *Food Science and Technology International* 7, 439–444.
- Mathiassen, J.R., Misimi, E., Bondø, M., Veliyulin, E., Østvik, S.O., 2011. Trends in application of imaging technologies to inspection of fish and fish products. *Trends in Food Science and Technology* 22, 257–275.

- Mendoza, F., Dejmek, P., Aguilera, J.M., 2006. Calibrated color measurements of agricultural foods using image analysis. *Postharvest Biology and Technology* 41, 285–295.
- OIV, 2013. Compendium of International Methods of Wine and Must Analysis. The International Organisation of Vine and Wine, Paris.
- Palus, H. 2006. Colorfulness of the image: definition, computation and properties. In: *Progress in Biomedical Optics and Imaging – Proceedings of SPIE* 6158.
- Rodríguez-Pulido, F.J., Ferrer-Gallego, R., González-Miret, M.L., Rivas-Gonzalo, J.C., Escribano-Bailón, M.T., Heredia, F.J., 2012. Preliminary study to determine the phenolic maturity stage of grape seeds by computer vision. *Analytica Chimica Acta* 732, 78–82.
- Romano, G., Argyropoulos, D., Nagle, M., Khan, M.T., Müller, J., 2012. Combination of digital images and laser light to predict moisture content and color of bell pepper simultaneously during drying. *Journal of Food Engineering* 109, 438–448.
- Salmerón, J.F., Gómez-Robledo, L., Carvajal, M.Á., Huertas, R., Moyano, M.J., Gordillo, B., et al., 2012. Measuring the colour of virgin olive oils in a new colour scale using a low-cost portable electronic device. *Journal of Food Engineering* 111, 247–254.
- The Mathworks, 2009. MATLAB (Version R2009b). The MathWorks Inc., Natick, Massachusetts.
- Urban, P., Berns, R.S., Grigaf, R.R. 2007. Color correction by considering the distribution of metamers within the mismatch gamut. In: *Final Program and Proceedings – IS and T/SID Color Imaging Conference*, pp. 222–226.
- Valous, N.A., Mendoza, F., Sun, D.W., Allen, P., 2009. Colour calibration of a laboratory computer vision system for quality evaluation of pre-sliced hams. *Meat Science* 81, 132–141.
- Wyszecki, G., 1954. A regular rhombohedral lattice sampling of munsell renotation space. *Journal of the Optical Society of America* 44, 725.
- Yager, R.R., Filev, D.P., 1994. Generation of fuzzy rules by mountain clustering. *Journal of Intelligent and Fuzzy Systems* 2, 209–219.
- Yam, K.L., Papadakis, S.E., 2004. A simple digital imaging method for measuring and analyzing color of food surfaces. *Journal of Food Engineering* 61, 137–142.
- Yin, H., Chai, Y., Yang, S.X., Mittal, G.S., 2011. Ripe tomato detection for robotic vision harvesting systems in greenhouses. *Transactions of the ASABE* 54, 1539–1546.
- Zheng, C., Sun, D.W., 2008. Image segmentation techniques. In: Sun, D.-W. (Ed.), *Computer Vision Technology for Food Quality Evaluation*. Academic Press, Amsterdam, pp. 37–56.



---

## CAPÍTULO 3

### RELACIÓN ENTRE EL COLOR Y LA COMPOSICIÓN FENÓLICA DE SEMILLAS DE UVA

Francisco José Rodríguez-Pulido, Francisco J. Heredia, Juan Manuel Zaldívar-Cruz, M. Lourdes González-Miret, (2012). *Application of Image Analysis to the Color-Phenolic Composition Relationships of Grape Seeds*. En: José Luis Caivano y María del Pilar Buera (Eds.) *Color in Food. Technological and Psychophysical Aspects*, 113 -120. CRC Press. ISBN 978-1-4398-7693-0

Francisco J. Rodríguez-Pulido, Raúl Ferrer-Gallego, M. Lourdes González-Miret, Julián Carlos Rivas-Gonzalo, María Teresa Escribano-Bailón, Francisco J. Heredia, (2012). *Preliminary study to determine the phenolic maturity stage of grape seeds by computer vision*, *Analytica Chimica Acta*, 732, 78-82, ISSN 0003-2670.

#### 3.1. Antecedentes

La colorimetría triestímulo es una ciencia extensamente aplicada en alimentos debido a la estrecha relación que existe entre el color y su composición química (Hutchings, 1999). Los métodos colorimétricos se han desarrollado más especialmente en campos donde se precisan análisis rápidos y automatizables que permitan el seguimiento de un determinado producto.

Las semillas de uva juegan un papel fundamental en el proceso de elaboración del vino y su calidad final. En este sentido, existen estudios que analizan en profundidad la contribución de los taninos de las semillas a las propiedades sensoriales de los vinos, como el cuerpo, la estructura, la astringencia y el color (Escribano-Bailón et al., 2001; Gawel, 1998). También existe un alto interés en este tipo de compuestos por sus propiedades antioxidantes y sus reconocidos efectos beneficiosos en la salud (Xia et al., 2010).

Son muchos los cambios físicos y químicos que se producen durante la maduración de las semillas, afectando al grado de polimerización, oxidación y extractabilidad de los polifenoles, y por ende, a sus propiedades organolépticas y a las propiedades del vino elaborado. Entre los compuestos fenólicos de interés, las catequinas y las proantocianidinas contribuyen directamente a la astringencia e indirectamente, a través de la interacción con antocianinas, al color de los vinos. La composición de las semillas de uva se caracteriza por la presencia de catequinas y

procianidinas, compuestas de (+)-catequina, (-)-epicatequina y unidades de (-)-epicatequin-3-O-galato, pues los compuestos galoilados siempre aparecen sobre la epicatequina, y nunca sobre la catequina. Entre los oligómeros, los dímeros son los más abundantes (Ferrer-Gallego et al., 2010).

En climas cálidos, la vid se enfrenta a primaveras secas y veranos calurosos. Este fenómeno hace que la pulpa alcance rápidamente altos niveles de azúcares mientras que los compuestos fenólicos de las semillas, resguardadas en el interior del fruto, no han alcanzado aún la madurez óptima. La versatilidad y la rapidez de las técnicas ópticas en general, y la visión artificial en particular, las convierten en una manera óptima para evaluar las semillas de uva, pudiendo establecer su estado de madurez a partir de los resultados del análisis de imagen mediante las herramientas estadísticas adecuadas. La monitorización de los cambios que suceden podría ser una importante herramienta para la toma de decisiones de índole agronómica, como por ejemplo la determinación del momento de vendimia.

### **3.2. Objetivos**

El objetivo general de este trabajo fue comprobar la aplicabilidad de la metodología puesta a punto en los estudios anteriores para predecir la composición de las semillas de uva. Los objetivos concretos planteados fueron:

- Evaluación de las semillas por análisis de imagen para la obtención de sus coordenadas colorimétricas y morfológicas.
- Evaluación química de las semillas de uva mediante análisis cromatográfico.
- Empleo de técnicas estadísticas univariantes y multivariantes para el estudio de la relación apariencia/composición y la designación de la etapa de madurez de las semillas de uva.

---

### 3.3. Parte experimental

El desarrollo completo del estudio está constituido por dos ensayos independientes. Se seleccionaron, para ello, diferentes muestras de semillas en diferentes estadios de madurez, de acuerdo a criterios químicos, de composición fenólica, y colorimétricos

#### 3.3.1. Estudio de la relación apariencia–composición mediante estadística univariante

El primer ensayo se realizó con semillas de uva tinta de la variedad Tempranillo de la Denominación de Origen Condado de Huelva, recogidas durante la campaña de 2009, en ocho muestreos, desde mediados de julio hasta el momento de la vendimia a finales de agosto. De una muestra bruta de 2 kg, se tomaron al azar cien bayas y se extrajeron sus semillas. Las imágenes fueron digitalizadas mediante el sistema DigiEye®. En cada imagen (muestra) se determinaron las coordenadas colorimétricas de los espacios RGB y CIELAB, así como las características morfológicas y de apariencia: largo, ancho y área de semilla, relación de aspecto, grado de redondez, heterogeneidad, grado de pardeamiento y la media de diferencias de color respecto al color medio (MCDM) (Berns, 2000). Respecto a las variables colorimétricas, se observó un descenso de la claridad y el croma de la semilla, que fue más leve en el caso del tono. Apenas se encontraron cambios notables en la morfología de la semilla en el intervalo de tiempo muestreado.

Como análisis químico de referencia en este ensayo se utilizó una modificación del método propuesto por Sandhu y Gu (2010). Éste utiliza cromatografía líquida de alta eficacia con detector de matriz de fotodiodos (HPLC-DAD). Para la identificación de los compuestos en los cromatogramas se compararon los tiempos de retención y los espectros con los de los patrones utilizados. Aunque se identificaron más compuestos, el ácido protocatéquico y la catequina mostraron una evolución más pronunciada. En la matriz de

correlaciones se pudo observar que prácticamente todas las relaciones fueron significativas ( $p < 0.05$ ). Además, no sólo correlacionaron las variables colorimétricas cuantitativas, sino también el tono ( $h_{ab}$ ), lo que pone de manifiesto cambios en la composición de la semilla desde un punto de vista cualitativo.

### 3.3.2. Caracterización del perfil fenólico y uso de técnicas multivariantes para la estimación de la etapa de madurez

El segundo ensayo incluido en el estudio se realizó con semillas de uva tinta de la variedad Graciano, recogidas en dos viñedos de la Denominación de Origen Calificada Rioja, durante la campaña de 2010, en seis muestreos desde el envero de la uva hasta sobremaduración. Cada muestra se separó en dos fracciones, enviadas a los dos laboratorios implicados en el estudio (Universidades de Salamanca y de Sevilla) para su análisis químico y de imagen, respectivamente.

En el Área de Nutrición y Bromatología de la Universidad de Salamanca, se analizaron los compuestos fenólicos mediante cromatografía líquida de alta eficacia con espectrometría de masas como sistema de detección (HPLC-MS) según el método de García-Marino et al. (2006). Se identificaron un total de veintidós compuestos fenólicos, que fueron agrupados en las siguientes familias: ácidos hidroxibenzoicos, monómeros de flavanoles, dímeros, trímeros y flavanoles galoilados. Al igual que en otros estudios publicados, el total de compuestos fenólicos descendió durante el periodo muestreado, y fueron los monómeros y los flavanoles galoilados los que sufrieron una disminución más acentuada. Ese descenso de los monómeros provocó que el cociente entre la concentración de monómeros y oligómeros (dímeros y trímeros) tendiera a disminuir durante la maduración. Así, ese cociente fue cercano a 1 en el envero y aproximadamente 0.6 en el momento de la vendimia, lo cual es coherente con las reacciones de polimerización que suceden durante la maduración. De todas las familias de compuestos fenólicos analizadas, los ácidos hidroxibenzoicos se encontraron en menor concentración y los que menos se modificaron durante el periodo considerado.

---

En el Laboratorio de Color y Calidad de Alimentos en la Universidad de Sevilla, la segunda fracción de muestra se empleó para el análisis de imagen. Se adquirieron las imágenes correspondientes a las semillas de cien bayas con el sistema DigiEye® y se procesaron según la metodología puesta a punto en los estudios anteriores. La evolución en los parámetros colorimétricos no fue tan notable como en el primer ensayo, donde el intervalo de tiempo considerado fue mayor. Debido a las condiciones climáticas de cada zona, el envero en las uvas de La Rioja se produce más tarde que en el Condado de Huelva, por lo que en la primera muestra que se tomó ya había comenzado el pardeamiento de las semillas y no se pudieron observar las primeras etapas del fenómeno, cuando los cambios son más acentuados. No obstante, se observó que las últimas etapas del pardeamiento se manifestaron en un descenso en los valores de las variables colorimétricas cuantitativas, claridad y croma. Además de las variables cuantitativas, también se modificó levemente el tono, lo cual indica que el fenómeno de pardeamiento también implica cambios desde un punto de vista cualitativo.

Como estudio preliminar, al igual que en el primer ensayo, se estudiaron las relaciones univariantes entre las concentraciones de los compuestos fenólicos, agrupados por familias, y las variables morfológicas y colorimétricas de las semillas medidas por análisis de imagen. Se encontraron altos coeficientes de determinación para algunas de esas relaciones, como entre monómeros y claridad ( $R^2=0.93$ ,  $p=0.008$ ), o flavanoles galoilados y croma ( $R^2=0.82$ ,  $p=0.047$ ). También se encontraron otras correlaciones estadísticamente significativas ( $p<0.05$ ) entre la composición fenólica y otras variables como heterogeneidad y longitud de semilla.

Para mejorar la predicción, y con la premisa de que unas variables podrían influir más que otras en la relación apariencia-composición, se aplicó una regresión múltiple por pasos (Forward Stepwise Multiple Regression) donde se intentó predecir la etapa de maduración en la que se encuentra la muestra, de composición química definida, a partir de las variables obtenidas por el análisis de imagen. El modelo de regresión incluyó secuencialmente la claridad, el croma, la

longitud de la semilla, el grado de redondez y la relación de aspecto, como variables independientes en la predicción. De ellas, fueron estadísticamente significativas ( $p < 0.05$ ) la claridad y la longitud de la semilla. Al evaluar la relación entre los valores predichos y los observados, se alcanzó un buen coeficiente de determinación ( $R^2 = 0.97$ ,  $p = 0.0089$ ). Este hecho, unido a la estrecha banda de regresión obtenida, indica una alta predictibilidad de la etapa de madurez a partir de los datos de análisis de imagen.

### 3.4. Conclusiones

- En general, se ha encontrado una disminución de la concentración de los compuestos identificados durante las etapas de maduración de acuerdo a las reacciones de polimerización y oxidación que se producen en este periodo. Esto corrobora los resultados obtenidos con anterioridad.
- Mediante la aplicación de técnicas multivariantes a los datos químicos y los obtenidos por análisis de imagen se verifica la existencia de una alta correlación entre la composición de las semillas de uva y sus coordenadas colorimétricas.
- Las características colorimétricas y morfológicas de las semillas pueden usarse para la predicción estimativa de algunos compuestos y familias de polifenoles en las semillas de uva.
- Aunque este método no pretende sustituir al análisis químico convencional, resulta una buena alternativa debido a la simplicidad, versatilidad y rapidez, así como al bajo coste que supone. Su uso implica ahorro de tiempo y de reactivos químicos, y podría ser útil en la industria vitivinícola para la toma de decisiones rápidas como la determinación del momento óptimo de vendimia.

# COLOR IN FOOD

---

Technological and  
Psychophysical Aspects

Edited by

José Luis Caivano

María del Pilar Buera



CRC Press

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

MATLAB® and Simulink® are trademarks of The MathWorks, Inc. and are used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® and Simulink® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® and Simulink® software.

CRC Press  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2012 by Taylor & Francis Group, LLC  
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper  
Version Date: 20120221

International Standard Book Number: 978-1-4398-7693-0 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Visit the Taylor & Francis Web site at**  
**<http://www.taylorandfrancis.com>**

**and the CRC Press Web site at**  
**<http://www.crcpress.com>**



CHAPTER 12  
Application of Image Analysis to  
the Color–Phenolic Composition  
Relationships of Grape Seeds

FRANCISCO JOSÉ RODRÍGUEZ-PULIDO, FRANCISCO  
J. HEREDIA, JUAN MANUEL ZALDÍVAR-CRUZ,  
and M. LOURDES GONZÁLEZ-MIRET

**Contents**

12.1 Introduction	113
12.2 Sampling	114
12.3 Image Analysis	114
12.4 Chemical Analysis	117
12.5 Conclusions	118
Acknowledgment	118
References	119

**12.1 Introduction**

The study of phenolic compounds in grape seeds of *Vitis vinifera* has achieved great interest in recent years. These compounds are responsible for structure and aroma of wine. Furthermore, these compounds are implicated in color of wine. Phenols such as catechin and its polymers are commonly present in grape seeds, which pass to the wine during fermentation process. Grape seeds are composed of water (25%–45%), polysaccharides (34%–36%), tannins (4%–10%), nitrogen compounds (4%–6.5%), minerals (2%–4%), and lipids (13%–20%). Although some of these phenolics are noncolored components, they influence the final color of red wine due to copigmentation reactions with anthocyanins (Boulton 2001). Copigmentation phenomenon consists of noncovalent molecular associations between anthocyanins and

other organic compounds such as flavonoids, yielding to changes or increments of the color intensity.

The composition of the seeds changes along the maturation until the grapes reach the ripeness, affecting the sensory properties of wine. In insufficiently mature grapes, seeds give aggressive and herbaceous flavor, yielding astringent taste to wine (Ristic and Iland 2005).

In warm climates, the wine is exposed to dry springs and hot summers. Due to this, the period between veraison and industrial maturity decreases. This phenomenon causes the grape pulp quickly reach the suitable sugar level to be fermented, while the phenolic compounds of the seeds have not yet reached optimum maturity. This will cause loss of color due to lack of phenols to stabilize the anthocyanic color in the wine. Currently, the moment of harvesting is determined based on chemical properties of the must, and the phenolic maturation of the seeds is not usually considered. The chemical changes occurring during the phenolic maturation induce changes of appearance of the seeds, modifying their color (from pale green to dark brown) as well as their shape.

On the other hand, colorimetry is widely used for evaluating the quality and composition of foods. Digital image analysis appears as a successful complement since it can be used to determine not only color but also other characteristics such as shape, texture, and homogeneity (Savakar et al. 2009, Zheng et al. 2006). In previous studies, we have determined the usefulness of the digital image analysis to assess the phenolic maturation evolution of the seeds in grapes for vinification. This has achieved interest in the industry. The morphological differences between varieties have been studied using general discriminant analysis models; these can classify grape seeds with high accuracy (Rodríguez-Pulido et al. 2010). This study has been continued using a deeper knowledge of the phenolic composition of the seeds and determining their relationships with appearance using digitization techniques.

## 12.2 Sampling

A red grape variety (*Tempranillo* cv.) grown in vineyards located in Condado de Huelva (Southwest of Spain) was sampled for this study twice a week from July until the harvest. A gross sample of 2 kg was divided into two fractions. One of these fractions was used for physicochemical analysis of the must, which determined average weight and berry size, pH, sugars, and total acidity. In the other fraction, 100 seeds were separated, cleaned, and imaged. Once the images were taken, the seeds were kept frozen until later HPLC analysis of phenols.

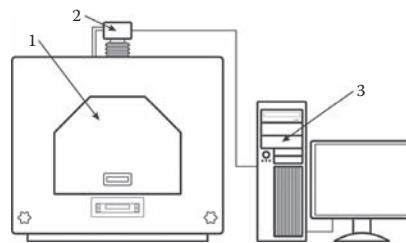
## 12.3 Image Analysis

Image analysis is framed by computer vision, a subfield of artificial intelligence, and its aim is to “teach” a computer to understand a scene and the

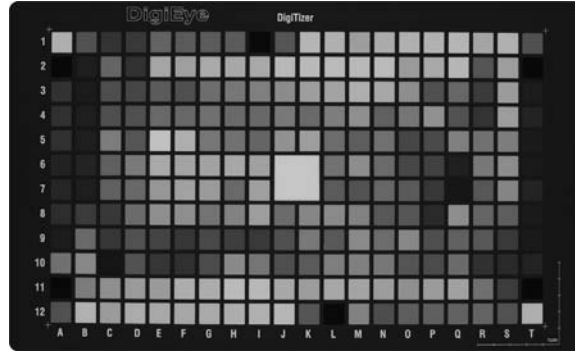
characteristics of an image. A computer vision system includes an illumination system, a charge-coupled device (CCD), a frame grabber that converts the analog image from the camera into a digitized one, and a computer with the suitable software for image processing and interpretation of results (Wang and Sun 2002). The digital camera receives images onto a CCD that registers them in gradations of three basic colors: red, green, and blue (RGB). This color space is widely used in electronic devices but it depends on the device where it is visualized. Thus, we cannot use this space for absolute measures of color. It is therefore necessary to convert between color spaces using a calibration. This requires control of illumination (CIE 2007, León et al. 2006). In this study, CIELAB space has been used. This was proposed by the International Commission on Illumination, CIE (1976), which defines any color by three attributes: lightness ( $L^*$ ), chroma ( $C_{ab}^*$ ), and hue ( $h_{ab}$ ). Besides color calibration, image analysis such as segmentation is an important feature of the analysis process. Segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something more meaningful and easier to analyze.

The main advantage of measuring color by digitalization derives from the fact that we can identify the seeds in the images by combining image analysis and tristimulus colorimetry technologies, and then morphological and colorimetric characteristics can be extracted from each one. Moreover, thanks to advances in computer science, this technique is becoming a powerful tool to test quality in the food industry (Brosnan and Sun 2004). Choosing the appropriate segmentation criteria, it is possible to analyze color and measure morphological properties in an automated manner (Zheng and Sun 2008).

In this chapter, the DigiEye<sup>®</sup> imaging system was used (Luo et al. 2001) (Figure 12.1). This includes an illumination box specially designed by VeriVide Ltd. (Leicester, United Kingdom) to illuminate the samples consistently in D65 illuminant and a digital camera connected to a computer (Figure 12.2). The camera parameters were set for all images with an exposure time of 1/15 s, aperture  $f/6.3$ , ISO 200 sensitivity, and were stored in TIF format with a size of  $3872 \times 2592$  pixels and a resolution of 96 PPI (pixels per inch).



**Figure 12.1** The DigiEye<sup>®</sup> system: (1) illumination box, (2) digital camera, and (3) computer.



**Figure 12.2** Chart of calibration of digital camera used by the DigiEye® system.

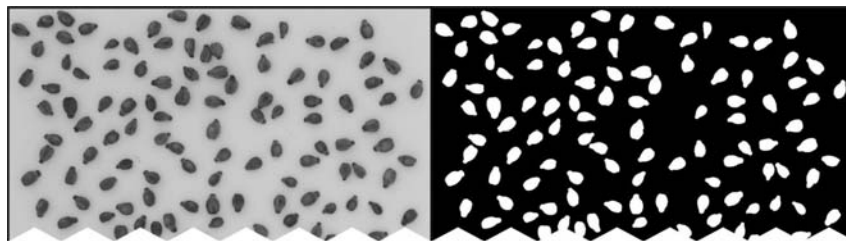
Using image analysis, morphological parameters and appearance were obtained as well as conversion from *RGB* values to CIELAB coordinates with original software DigiFood® (Heredia et al. 2006). Segmentation criteria were based on HSI (hue/saturation/intensity) color space and morphological features of seeds, such as area and size. HSI color space stems from *RGB* but its attributes have the same psychophysical meaning as that of the angular coordinates of CIELAB space. The colorimetric segmentation criteria were hues (*H*) between 0 and 255 (all hues), saturation (*S*) between 0 and 180, and intensity (*I*) between 0 and 140. Furthermore, these objects must have an area between 6 and 35 mm<sup>2</sup> and length not greater than 15 mm. These dimensions are consistent with the sizes of the seeds. For this purpose and in order to establish a correspondence between both magnitudes, the software performed a spatial calibration to convert pixels to millimeters.

From segmented images, the following parameters were measured:

- Length, width, and area
- Aspect: Ratio between major and minor axes of the ellipse equivalent to seed
- Roundness =  $\frac{\text{Perimeter}^2}{4 \times \pi \times \text{area}}$
- Heterogeneity: Fraction of pixels that deviate >10% from the average intensity
- Browning level: Fraction of pixels having  $L^*$  value is lower than 50 units
- Mean color difference from the mean (MCDM) (Berns 2000):

$$\text{MCDM} = \frac{\sum_{i=1}^N \left[ \left( L_i^* - \bar{L}^* \right)^2 + \left( a_i^* - \bar{a}^* \right)^2 + \left( b_i^* - \bar{b}^* \right)^2 \right]^{1/2}}{N}.$$

- RGB color



**Figure 12.3** Image of grape seeds and resulting mask after apply segmentation process.

**Table 12.1** Colorimetric Values for Grape Seeds during Maturation

Sample	Date	$L^*$	$a^*$	$b^*$	$C_{ab}^*$	$h_{ab}$
1	Jul 20	57.97	10.08	29.10	31.09	69.65
2	Jul 23	54.97	10.36	27.33	29.60	67.70
3	Jul 27	55.97	11.24	27.55	29.98	67.25
5	Aug 3	51.49	11.17	23.92	26.60	64.24
6	Aug 6	51.01	11.72	24.05	26.95	62.89
7	Aug 10	50.92	12.02	23.74	26.79	62.26
8	Aug 13	50.15	11.11	21.41	24.31	61.58

Once the images were acquired and the criteria of segmentation were applied, all seeds were recognized by the software (Figure 12.3).

Table 12.1 shows the CIELAB colorimetric parameters at each stage of maturation. Lightness, chroma, and hue all decreased with time due to the browning of the seed.

#### 12.4 Chemical Analysis

For the analysis of phenols, a modification of the method proposed by Sandhu and Gu (2010) was applied. 0.5 g of freeze-dried seeds were milled and extracted with 10 mL of acetone/water/acetic acid (70:29.3:0.3 v/v) solvent. After this, acetone from extraction was removed with the help of a vacuum concentrator. The remaining fraction was redissolved in an aqueous solution of formic acid. This solution was filtered in nylon and was encapsulated for the injection. Chromatographic analysis was performed on an Agilent 1200 series HPLC system equipped with a UV-visible diode array detector.

The phenolic profile of seeds was analyzed. For identification of these compounds, retention times and spectra were compared. Although other compounds were identified, Table 12.2 shows only those compounds whose concentration changed.

For the study of relationship between chemical composition and color in seeds, linear and multiple regression models were applied. As shown in

**Table 12.2** Content of Phenolic Compounds in Grape Seeds

Date	Protocatechuic Acid (mg/kg)	Catechin (mg/kg)
Jul 20	8	867
Jul 23	8	985
Jul 27	12	674
Aug 3	26	504
Aug 6	20	604
Aug 10	38	443
Aug 13	24	491

**Table 12.3** Correlation Matrix between Colorimetric and Chemical Values

	$L^*$	$a^*$	$b^*$	$C_{ab}^*$	$h_{ab}$
Protocatechuic acid (mg/kg)	<b>-0.82</b>	<b>0.81</b>	<b>-0.78</b>	-0.75	<b>-0.84</b>
Catechin (mg/kg)	<b>0.80</b>	<b>-0.83</b>	<b>0.82</b>	<b>0.80</b>	<b>0.86</b>

Note: Results in bold are significant at  $p < 0.05$ .

Table 12.3, the quantitative colorimetric variables correlated with not only some phenolic compounds, but also the hue ( $h_{ab}$ ), demonstrating their involvement in browning phenomenon from a qualitative point of view.

## 12.5 Conclusions

Features in images such as color, size, and shape can be extensively applied to seed maturity evaluation. The results suggest that image analysis has good potential to be used for color and appearance measurements in grape seeds. Moreover, as has been demonstrated in other food products, there is a high correlation between the chemical and food color. For this reason, this technique provides a fast, nondestructive, and low-cost alternative in the winery industry to conventional chemical analysis. The digital imaging method provides measurements and analyses of the color of food surfaces that are adequate for food science research. While it is not yet a replacement for sophisticated color measurement instruments, it is an attractive alternative due to its simplicity, versatility, and low cost.

## Acknowledgment

This work was supported by funding from the Ministerio de Ciencia e Innovación, Gobierno de España by the project AGL2008-05569-C02-02.

## References

- Berns, R. 2000. *Billmeyer and Saltzman's Principles of Color Technology*. New York: John Wiley & Sons.
- Boulton, R. 2001. The copigmentation of anthocyanins and its role in the color of red wine: A critical review. *American Journal of Enology and Viticulture* 52(2): 67–87.
- Brosnan, T. and D. W. Sun. 2004. Improving quality inspection of food products by computer vision—A review. *Journal of Food Engineering* 61(1): 3–16.
- CIE (Commission Internationale de l'Éclairage). 1976. *Recommendations on Uniform Color Spaces, Color-Difference Equations, Psychometric Color Terms*. Vienna, Austria: CIE Central Bureau.
- CIE (Commission Internationale de l'Éclairage). 2007. *Standard Illuminants for Colorimetry*. Vienna: CIE Central Bureau. ISO 11664-2:2007.
- Heredia, F. J., M. L. González-Miret, C. Álvarez, and A. Ramírez. 2006. DigiFood. Registration No. SE-01298.
- León, K., D. Mery, F. Pedreschi, and J. León. 2006. Color measurement in  $L^*a^*b^*$  units from RGB digital images. *Food Research International* 39(10): 1084–1091.
- Luo, M. R., G. H. Cui, and C. Li. 2001. Apparatus and method for measuring colour (DigiEye System), Derby University Enterprises Limited. British Patent 0124683.4.
- Ristic, R. and P. G. Iland. 2005. Relationships between seed and berry development of *Vitis vinifera* L. cv. Shiraz: Developmental changes in seed morphology and phenolic composition. *Australian Journal of Grape and Wine Research* 11(1): 43–58.
- Rodríguez-Pulido, F. J., L. Gómez-Robledo, M. L. González-Miret, and F. J. Heredia. 2010. Seguimiento de la maduración de variedades de uva mediante colorimetría y análisis de imagen de semillas. In *IX Congreso Nacional del Color*. Alicante, Spain.
- Sandhu, A. K. and L. Gu. 2010. Antioxidant capacity, phenolic content, and profiling of phenolic compounds in the seeds, skin, and pulp of *Vitis rotundifolia* (Muscadine Grapes) as determined by HPLC-DAD-ESI-MSn. *Journal of Agricultural and Food Chemistry* 58(8): 4681–4692.
- Savakar, D., G. Anami, and S. Basavaraj. 2009. Classification of food grains, fruits and flowers using machine vision. *International Journal of Food Engineering* 5(4): art. 14.
- Wang, H. H. and D. W. Sun. 2002. Correlation between cheese meltability determined with a computer vision method and with Arnott and Schreiber tests. *Journal of Food Science* 67(2): 745–749.
- Zheng, C. and D. W. Sun. 2008. Object measurement methods. In *Computer Vision Technology for Food Quality Evaluation*. Amsterdam, the Netherlands: Academic Press.
- Zheng, C., D. W. Sun, and L. Zheng. 2006. Recent developments and applications of image features for food quality evaluation and inspection—A review. *Trends in Food Science & Technology* 17(12): 642–655.







## Preliminary study to determine the phenolic maturity stage of grape seeds by computer vision

Francisco J. Rodríguez-Pulido<sup>a</sup>, Raúl Ferrer-Gallego<sup>b</sup>, M. Lourdes González-Miret<sup>a</sup>, Julián Carlos Rivas-Gonzalo<sup>b</sup>, María Teresa Escribano-Bailón<sup>b</sup>, Francisco J. Heredia<sup>a,\*</sup>

<sup>a</sup> Food Colour & Quality Lab., Dept. Nutrition & Food Science, Facultad de Farmacia, Universidad de Sevilla, 41012 Sevilla, Spain

<sup>b</sup> Grupo de Investigación en Polifenoles, Unidad de Nutrición y Bromatología, Facultad de Farmacia, Universidad de Salamanca, Campus Miguel de Unamuno, E 37007 Salamanca, Spain

### ARTICLE INFO

#### Article history:

Received 1 September 2011

Received in revised form

29 December 2011

Accepted 4 January 2012

Available online 10 January 2012

#### Keywords:

Image analysis

Colour

*Vitis vinifera*

Grape seeds

Phenolics

### ABSTRACT

The applications of computer vision technology for acquiring and analysing images have been extended to the quality evaluation in food industry. This technique involves great advantages for the objective, rapid, non-contact and automated quality inspection and control. The aim of this work was to evaluate the potential of the computer vision to determine the phenolic maturity stage of grape seeds. Up to 21 phenolic compounds were determined by HPLC–DAD–MS in order to obtain reference values to develop the model. The CIELAB parameters, area, aspect, roundness, length, width and heterogeneity of seeds were analysed using a DigiEye<sup>®</sup> system. The technique reported in this work can be a good and rapid tool for taking decisions at harvest time. Notwithstanding, a comprehensive study should be made in order to develop more robust models.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

The phenolic composition of grape seeds plays an important role on the red wine quality. In this sense, the contribution of seed tannins to several sensorial attributes of wines, such as body, structure, bitterness, astringency and colour are well known [1,2]. Also, the presence of phenolic compounds in wine has been the subject of great interest due to their potentially beneficial effects on human health [3].

Many physical and biochemical changes take place during grape seeds maturity which affect their characteristics like astringency, tannic intensity, colour and morphology [4,5]. Quantity, structure and degree of polymerization of seed tannins are modified during ripening development [6,7]. Monitoring these changes could be an important tool for deciding the optimal moment for harvesting, but this would require to have rapid analytical methods involving minimal or no sample preparation.

Colour and appearance are closely related to the chemical composition and sensory properties of food. Usually, colour is determined by tristimulus colorimetry and expressed in terms of  $L^*a^*b^*$  variables, corresponding to the uniform colour space CIELAB.

The CIELAB colour space (1976  $L^*a^*b^*$  colour space) is an international standard for colour measurements, recommended by the Commission International d'Eclairage (CIE) in 1976. Within CIELAB, a psychometric index of lightness ( $L^*$ ) and two colour coordinates ( $a^*$  and  $b^*$ ) are defined.  $L^*$  is a qualitative attribute of relative luminosity, which is the property according to which each colour can be considered as equivalent to a member of the grey scale, ranging between black ( $L^* = 0$ ) and white ( $L^* = 100$ ). Coordinate  $a^*$  takes positive values for reddish colours and negative values for greenish ones, whereas  $b^*$  takes positive values for yellowish colours and negative values for bluish ones [8]. From these Cartesian coordinates, two angular parameters can be defined: chroma ( $C^*_{ab}$ ) and hue ( $h_{ab}$ ) [9]. Hue angle ( $h_{ab}$ ) is the qualitative attribute of colour (blue, purple, green, . . .), and chroma ( $C^*_{ab}$ ) is considered the quantitative attribute of colourfulness, allowing assessing the degree of difference of any given hue relative to a grey colour with the same lightness.

The CIELAB colour space is considered as uniform by the CIE, i.e., the Euclidean distance between two different colour points corresponds approximately to the colour difference perceived by the human eye.

For colour measurement, different kinds of instruments such as colorimeters, spectrophotometers and spectroradiometers can be used. However, in the case of non-homogeneous colour, these instruments require homogenizing the sample to achieve uniform

\* Corresponding author. Tel.: +34 954556495; fax: +34 954557017.

E-mail address: [heredia@us.es](mailto:heredia@us.es) (F.J. Heredia).

colour. This fact makes complicated and tedious measuring colour of heterogeneous stuff or small objects, such as grape seeds.

Computer vision provides fast, accurate, direct and non-destructive analysis that this case needs. Moreover, not only colour but also other characteristics such as shape, texture and homogeneity can be determined [10,11].

A computer vision system includes an illumination system, a charge-coupled device (CCD), a frame grabber which converts the analogue image from the camera into a digitized one, and a computer with the suitable software for image processing and interpretation of results [12]. The digital camera receives images onto a CCD that registers them in gradations of three basic colours: red, green and blue (RGB colour space). This colour space is widely used in electronics devices but it depends on the device where it is visualized. So, we cannot use this space for absolute measures; it is necessary to transform the data from RGB to CIELAB colour spaces. This transformation requires calibration and depends on the illumination when images are taken.

In the image analysis perform, it is indispensable to recognize objects into the image before analyse them. For this purpose, the representation of the image is changed into something more meaningful and easier to analyse [13]. This process is known as segmentation and refers to the partition of a digital image into multiple sets of pixels. Once this recognition is done, the morphology of the objects are analysed and colour coordinates are calculated from the pixels that they are composed of. Frequently, HSI (hue, saturation and intensity) space is used for the segmentation process [14]. HSI colour space is easier to understand than RGB due to its likeness with the  $L^*$ ,  $C^*_{ab}$  and  $h_{ab}$  stimuli, which are more related to the visual appreciation of colour.

The aim of this study was to find relationships between chemical (phenolic composition) and appearance (colour and morphological) data of grape seeds which can yield to the assessment of the phenolic maturation stage based on parameters obtained by image analysis.

## 2. Materials and methods

### 2.1. Samples

*Vitis vinifera* L. cv. Graciano red grape berries grown on Rioja Designation of Origin (Spain) were collected at two vineyards in 2010, at six different stages during the berry maturation (from veraison until over-ripeness). Sampling was carried out as follows: 150 berries were collected from both sides of vines in a row within the vineyard. Edge rows and the first two vines in a row were avoided. In order to have representative samples, berries were collected from the top, middle and bottom of the clusters from different vines along the vineyard. The samples (whole grapes) were immediately frozen and stored at  $-20^\circ\text{C}$  until the analyses were performed.

### 2.2. Phenolic extraction and analysis

Phenolic extraction was carried out as described in García-Marino et al. [15]. For extraction, grape seeds were separated manually and freeze-dried. After this, samples were ground to obtain homogeneous powder. 2 g of homogeneous powder was extracted with twenty millilitres of methanol:water (75:25). The extraction process was repeated twice more. The methanolic extracts were combined and concentrated at low pressure until aqueous extract was obtained. The extracts were filtered through a  $0.45\ \mu\text{m}$  filter and injected directly into the chromatographic system. All analyses were performed in triplicate.

Liquid chromatography–diode array detection–mass spectrometry (LC–DAD–MS) analysis was performed using an Agilent

Technologies 1200 series liquid chromatograph (Palo Alto, CA, USA). A Spherisorb® S3 ODS-2 reverse phase  $3\ \mu\text{m}$  particle size C18 column  $150\ \text{mm} \times 4.6\ \text{mm}$  i.d. (Waters, Ireland) thermostated at  $25^\circ\text{C}$  was used. The solvents used were (A) 2.5% acetic acid, (B) 2.5% acetic acid/acetonitrile (90:10, v:v), and (C) HPLC-grade acetonitrile, establishing the following gradient: from 0 to 100% B for 5 min, from 0 to 15% C for 25 min, from 15 to 50% C for 5 min, and isocratic 50% C for 5 min, at a flow rate of  $0.5\ \text{mL min}^{-1}$ . Detection was carried out at 280 nm as preferred wavelength. The mass spectrometer (MS) was connected to HPLC system via the DAD cell outlet. MS detection was performed with API 3200 Qtrap (Applied Biosystems, Darmstadt, Germany) equipped with an electrospray ionization (ESI) source and a triple quadrupole-ion trap mass analyzer, controlled by the Analyst 5.1 software.

Quantification was performed by HPLC–DAD using calibration curves of (+)-catechin, purchased from Sigma–Aldrich (St. Louis, USA), and of procyanidin standards previously obtained in our laboratory [16,17].

### 2.3. Image acquisition

The DigiEye® imaging system based upon the calibrated digital camera was used [18]. This device consists on a closed illumination box, specially designed by VeriVide Ltd. (Leicester, UK), and a digital camera (10.2-megapixel Nikon® D80 with Nikkor® 35 mm f/2D objective) connected to a computer (Pentium IV 3.00 GHz processor) via USB. The cabinet is equipped with two CIE D65 standard illuminant emulators which allow illuminating the samples consistently under stable lighting conditions. For obtaining morphological and appearance parameters, as well as the CIELAB coordinates from RGB colour space data, the software DigiFood® [19] was used.

## 3. Results and discussion

### 3.1. Chemical analysis

Based on the chromatographic and spectral features, twenty-one phenolic compounds were identified. These were quantified and grouped into hydroxybenzoic acids, flavanol monomers (including catechin and epicatechin), dimers of procyanidins, trimers of procyanidins and galloylated flavanols (epicatechin 3-O-gallate and galloylated procyanidins). Fig. 1 shows the evolution during ripening of the studied phenolic compounds. These compounds correspond to the main phenolic compounds present in the chromatograms of the seed extracts.

Total phenolic content, defined as the sum of the content of the phenolic compounds identified, decrease during ripening, in agreement with other works [4,6,20–22]. Monomers and galloylated flavanols undergo the sharpest drop. Near maturity, the content of galloylated compounds is close to that found by other authors for this variety (24% of total) [4,23]. The pronounced decrease of the flavanol monomers provokes that the ratio of the concentration of monomers to oligomers (dimers and trimers) tends to diminish during ripening. Thus, values close to 1 can be calculated at veraison while ca. 0.6 is obtained at harvest. This is consistent with possible polymerization reactions occurring during maturation.

Hydroxybenzoic acids are present in seeds in low amounts (from  $0.4$  to  $0.5\ \text{mg g}^{-1}$ ) and their concentration do not suffer as noticeable changes during ripening as flavanols.

### 3.2. Image analysis

Images of the seeds corresponding to 100 berries per sample were acquired by using the DigiEye system. For seed recognition

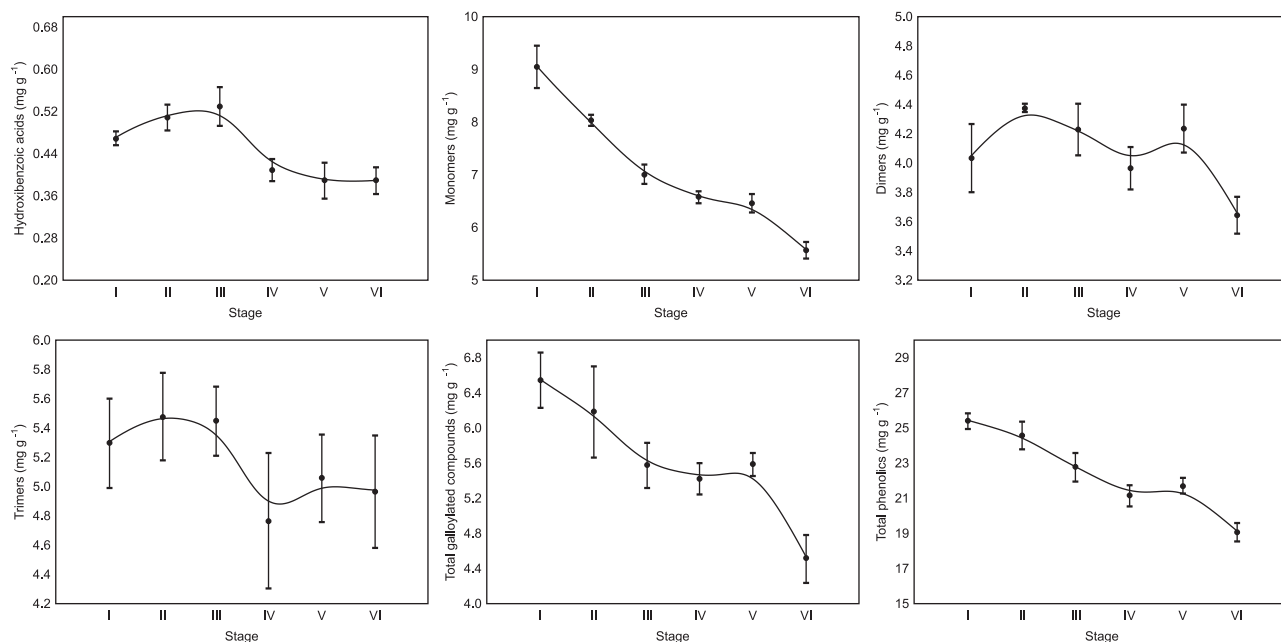


Fig. 1. Contents of the different polyphenols divided in groups at each stage of maturation.

within the images, the following selection criteria were established: (a) regions having values of saturation (S) lower than 180 units, and intensity (I) lower than 140 units, (b) objects having area values between 6 and 35 mm<sup>2</sup>, and length values less than 15 mm. By applying these recognition criteria, all the seeds were included for every image, preventing the recognition of undesirable objects in the digital image.

### 3.2.1. Colour

RGB colour was obtained from raw images, which involves minimally processed data from the CCD device. Then, CIELAB colour coordinates were obtained from the RGB values, taking into account the camera calibration. The colorimetric data of digital images (Table 1) entails high standard deviation values due to the chromatic heterogeneity of the samples.

The phenomenon of browning appears as a decrease of the quantitative colorimetric parameters ( $L^*$  and  $C^*_{ab}$ ). This fact is evidenced in a darkening and a lack of chromaticity of the final samples of grape seeds. Browning advances while chemical composition changes, so it suggests a connection between these two phenomena.

Not only the quantitative psychophysical colorimetric parameters ( $L^*$  and  $C^*_{ab}$ ) correlated with some phenolic compounds, but also the hue ( $h_{ab}$ ), demonstrating that the browning phenomenon implies changes also from a qualitative point of view.

Additionally, chromatic heterogeneity was measured, considering being this variable the fraction of pixels that deviate more than 10% from the average intensity. Chromatic heterogeneity increases in the early stages of maturation and decreases in the final stages,

because browning process does not occur in a homogeneous manner. The seeds become dark brown bit by bit while the light green areas disappear in the same way.

### 3.2.2. Morphological variables

The spatial unit in imaging is the pixel, so for calculating the morphological data from the images we need to perform the spatial calibration to find the correspondence between pixel and millimetre. With this purpose, the image of a calliper was taken in the same way than the seeds images.

The size of the seeds was measured in two directions: diameter along major axis of the seed (*length*) and diameter along the axis perpendicular to the major axis (*width*). In order to measure the elongation degree of each seed, the *aspect ratio* was calculated which is the ratio between major and minor axis of the ellipse equivalent to the seed. Also the *roundness* value was calculated by using the following dimensionless equation:

$$\text{Roundness} = \frac{\text{perimeter}^2}{4 \times \pi \times \text{area}}$$

in such a way that a perfect circle has a roundness value of one and it becomes higher while the shape elongates.

The *area* of each pixel was calculated by the spatial calibration. Thus, the area of each seed was calculated by pixel count.

### 3.3. Statistical treatment

Once the chemical and image analysis were applied for all the samples, direct univariate relationships (linear regression models)

Table 1  
CIELAB values of the grape seeds at each stage of maturation. Mean  $\pm$  Std. Dev. (CIELAB units.)

	Stage					
	I	II	III	IV	V	VI
$L^*$	45.1 $\pm$ 6.7	43.4 $\pm$ 6.8	41.5 $\pm$ 7.0	40.3 $\pm$ 6.9	42.3 $\pm$ 6.6	40 $\pm$ 7.3
$a^*$	7.6 $\pm$ 3.1	7.6 $\pm$ 2.9	7.6 $\pm$ 2.9	8.2 $\pm$ 3.0	6.4 $\pm$ 2.7	8.4 $\pm$ 3.0
$b^*$	23.8 $\pm$ 6.8	22.1 $\pm$ 6.5	21.9 $\pm$ 6.4	21.9 $\pm$ 6.6	21.5 $\pm$ 6.3	21.4 $\pm$ 6.7
$C^*_{ab}$	25.2 $\pm$ 6.8	23.6 $\pm$ 6.4	23.4 $\pm$ 6.3	23.6 $\pm$ 6.6	22.6 $\pm$ 6.2	23.2 $\pm$ 6.7
$h_{ab}$	71.9 $\pm$ 7.6	70.3 $\pm$ 8.2	70.2 $\pm$ 9.3	68.8 $\pm$ 7.7	72.5 $\pm$ 8.5	68.0 $\pm$ 11.2

**Table 2**Correlation matrix ( $R^2$ ) between chemical composition and parameters obtained by image analysis in each stage. Results in bold are significant at  $\alpha = 0.05$ .

	Hydroxybenzoic acids	Monomers	Dimers	Trimers	Galloylated compounds
$L^*$	0.54	<b>0.93</b>	0.39	0.35	0.80
$a^*$	-0.38	-0.18	0.16	0.41	0.00
$b^*$	0.27	<b>0.91</b>	0.38	0.46	<b>0.85</b>
$C^*_{ab}$	0.17	<b>0.83</b>	0.41	0.55	<b>0.82</b>
$h_{ab}$	0.45	0.57	0.01	-0.18	0.38
Area	-0.52	-0.32	<b>-0.88</b>	-0.64	-0.55
Aspect	-0.42	-0.20	0.03	-0.26	-0.03
Roundness	0.00	-0.20	0.31	0.08	0.01
Length	<b>-0.91</b>	-0.59	-0.77	-0.81	-0.61
Width	-0.10	-0.13	-0.57	-0.23	-0.36
Heterogeneity	0.72	0.42	0.77	<b>0.94</b>	0.53

**Table 3**Results of forward stepwise Multiple Regression. Results in bold are significant at  $\alpha = 0.05$ .

	Beta	Std. Err. of Beta	B	Std. Err. of B	p-level
Intercept			114.399	18.817	0.0089
$L^*$	-0.816	0.196	-0.680	0.164	<b>0.0254</b>
Length	0.513	0.096	9.840	1.839	<b>0.0128</b>
Aspect	-0.316	0.131	-11.836	4.918	0.0953
Roundness	-0.216	0.127	-15.160	8.901	0.1871
$C^*_{ab}$	-0.182	0.156	-0.329	0.281	0.3271

between the chemical information and the parameters obtained by image analysis were studied. As it can be observed in Table 2, high correlations have been found between colour and phenolic composition, such as monomers and lightness ( $L^*$ ) ( $R^2 = 0.93$ ,  $p = 0.008$ ) or galloylated compounds and chroma ( $C^*_{ab}$ ) ( $R^2 = 0.82$ ,  $p = 0.047$ ). Moreover, some high correlations were also found with morphological and appearance data, such as trimers and heterogeneity ( $R^2 = 0.94$ ,  $p = 0.005$ ) or hydroxybenzoic acids and length ( $R^2 = -0.91$ ,  $p = 0.011$ ). These considerable correlations with some morphological parameters were the reason why they were also included in further multivariate analysis.

Thus, a forward stepwise multiple regression was applied for estimating the maturity stage of grape seeds from the values obtained by image analysis. The multiple regression models included sequentially lightness ( $L^*$ ), chroma ( $C^*_{ab}$ ), seed length, roundness and aspect ratio as independent variables in the prediction, being lightness ( $L^*$ ) and length the variables which influenced significantly ( $p < 0.05$ ) (Table 3). Fig. 2 shows the scatterplot of observed vs. predicted values of maturity stage, showing linear fit with quite good correlation coefficient ( $R^2 = 0.97$ ,  $p = 0.0089$ ). This fact, together to the narrow regression bands (the confidence level of the bands represents the probability that additional sample

points will fall within the prediction interval) indicate the high predictability of the maturity stage based on image analysis data.

#### 4. Conclusions

Twenty one identified phenolic compounds decreased their concentration along the different grape seeds maturation stages studied, related to polymerization reactions occurring along this period, and the colorimetric and morphological parameters analysed by image analysis also changed.

In some cases, single linear regression models show that there are good relationships between chemical and appearance data. Moreover, it is possible to estimate the maturity stage of grape seeds by applying a forward stepwise multiple regression models to these data, being achieved a correlation coefficient of  $R^2 = 0.97$  for predicting the maturity stage.

The results suggest that image analysis has good potential to be used for colour and appearance measurements in grape seeds. It is not yet a substitute for conventional chemical analysis but it is an attractive alternative due to its simplicity, versatility, and low-cost.

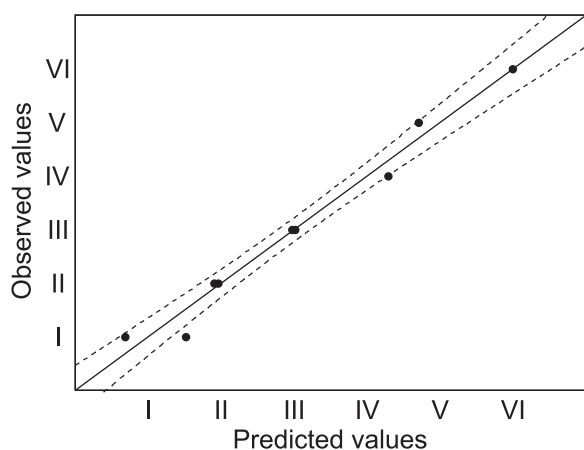
Although the suggested method should be applied to other grape varieties to study its robustness, it is a quick and objective way to estimate the maturity stage. Controlling the variable affecting, it could become a way to assess the chemical characteristics of grape seeds during maturation, saving time and chemical reagents and thus allowing the winery to take quick decisions, for example in determining the moment of harvesting.

#### Acknowledgements

We thank the Ministry of Science and Innovation of Spain the financial support: Projects AGL2008-05569-C02-01 and AGL2008-05569-C02-02, and fellowship BES-2009-025429.

#### References

- [1] T. Escribano-Bailón, M. Álvarez-García, J.C. Rivas-Gonzalo, F.J. Heredia, C. Santos-Buelga, J. Agric. Food Chem. 49 (2001) 1213–1217.
- [2] R. Gawel, J. Aust. Grape Wine Res. 4 (1998) 74–95.
- [3] E.Q. Xia, G.F. Deng, Y.J. Guo, H.B. Li, Int. J. Mol. Sci. 11 (2010) 622–646.
- [4] R. Ferrer-Gallego, M. García-Marino, J.M. Hernández-Hierro, J.C. Rivas-Gonzalo, M.T. Escribano-Bailón, Anal. Chim. Acta 660 (2010) 22–28.
- [5] R. Ristic, P.G. Iland, Aust. J. Grape Wine Res. 11 (2005) 43–58.



**Fig. 2.** Scatterplot of observed vs. predicted values of stage of maturity. Dashed lines shows confidence bands at (95%).

- [6] V.A.P. De Freitas, Y. Glories, A. Monique, *Am. J. Enol. Vitic.* 51 (2000) 397–403.
- [7] J.A. Kennedy, Y. Hayasaka, S. Vidal, E.J. Waters, G.P. Jones, *J. Agric. Food Chem.* 49 (2001) 5348–5355.
- [8] K.L. Yam, S.E. Papadakis, *J. Food Eng.* 61 (2004) 137–142.
- [9] Commission Internationale de l'Éclairage (CIE), Technical Report, Colorimetry, CIE Central Bureau, Vienna 2004.
- [10] D.G. Savakar, S.B. Anami, *Int. J. Food Eng.* 5 (2009) (Art. 14).
- [11] C. Zheng, D.W. Sun, in: D.W. Sun (Ed.), *Computer Vision Technology for Food Quality Evaluation*, Academic Press, Amsterdam, 2008, pp. 57–80.
- [12] H.H. Wang, D.W. Sun, *J. Food Sci.* 67 (2002) 745–749.
- [13] C. Zheng, D.W. Sun, in: D.W. Sun (Ed.), *Computer Vision Technology for Food Quality Evaluation*, Academic Press, Amsterdam, 2008, pp. 37–56.
- [14] K. Jack, *Digital Video*, D.S.P. Newnes, Burlington (2008) 15–29.
- [15] M. García-Marino, J.C. Rivas-Gonzalo, E. Ibáñez, C. García-Moreno, *Anal. Chim. Acta* 563 (2006) 44–50.
- [16] S. de Pascual-Teresa, J.C. Rivas-Gonzalo, C. Santos-Buelga, *Int. J. Food Sci. Technol.* 35 (2000) 33–40.
- [17] M.T. Escribano-Bailón, Y. Gutiérrez-Fernández, J.C. Rivas-Gonzalo, C. Santos-Buelga, *J. Agric. Food Chem.* 40 (1992) 1794–1799.
- [18] M.R. Luo, F.H. Cui, C. Li, British Patent Entitled Apparatus and Method for Measuring Colour (DigiEye System), Derby University Enterprises Limited, 2001 (0124683.4).
- [19] F.J. Heredia, M.L. Gonzalez-Miret, C. Álvarez, A. Ramírez, *DigiFood* [No. SE-01298], 2006.
- [20] J.A. Kennedy, M.A. Matthews, A.L. Waterhouse, *Phytochemistry* 55 (2000) 77–85.
- [21] M.O. Downey, J.S. Harvey, S.P. Robinson, *Aust. J. Grape Wine Res.* 9 (2003) 15–27.
- [22] A.M. Jordao, J.M. Ricardo Da Silva, O. Laureano, *Am. J. Enol. Vitic.* 52 (2001) 230–234.
- [23] V. Núñez, C. Gómez-Cordovés, B. Bartolomé, Y.J. Hong, A.E. Mitchell, *J. Sci. Food Agric.* 86 (2006) 915–921.



---

## CAPÍTULO 4

### APLICACIÓN DE LAS TÉCNICAS HIPERESPECTRALES A LAS SEMILLAS DE UVA. ESTABLECIMIENTO DE LA METODOLOGÍA Y DISCRIMINACIÓN VARIETAL

Francisco J. Rodríguez-Pulido, Douglas F. Barbin, Da-Wen Sun, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia, (2013). *Grape seed characterization by NIR hyperspectral imaging*, *Postharvest Biology and Technology*, 76, 74-82, ISSN 0925-5214.

#### 4.1. Antecedentes

A diferencia de los sistemas de imágenes basados en filtros, como son las cámaras a color convencionales, las cámaras hiperespectrales proporcionan una alta resolución en diferentes intervalos del espectro. La imposibilidad de abarcar grandes intervalos proviene de las limitaciones en los detectores, y dentro de los posibles intervalos, el infrarrojo cercano (NIR, 750-3000 nm) ha conducido a los mejores resultados en ciencia de los alimentos (Bruno & Svoronos, 2005; Osborne et al., 1993).

La importancia del salto de las técnicas de imagen a color a las técnicas hiperespectrales en el NIR estriba, principalmente, en la detección de características que no podrían ser evaluadas mediante el espectro visible (Choudhary et al., 2009; Lacar et al., 2001). Además, la medida a múltiples longitudes de onda permite la detección simultánea de varios constituyentes de una muestra, ya que las moléculas orgánicas tienen patrones de absorbanza específicos que pueden ser utilizados para la caracterización de la composición química (Williams & Norris, 2001). Las técnicas de imágenes hiperespectrales ya han sido utilizadas con éxito en productos agroalimentarios (Baiano et al., 2012; Baye et al., 2006; Cubero et al., 2011; Lorente et al., 2012; Shahin & Symons, 2011), pero aún no se han aplicado en la determinación de características fisicoquímicas en semillas de uva. Al igual que la mayoría de las técnicas de visión artificial, los sistemas hiperespectrales actúan solo en las capas exteriores de los materiales. No obstante, es en estas zonas donde se concentran mayoritariamente los compuestos fenólicos de interés en las semillas de uva (Thorngate & Singleton, 1994).

Las imágenes hiperespectrales o hipercubos están formadas por una batería de imágenes bidimensionales, donde cada una de ellas muestra en escala de grises la reflectividad a una longitud de onda. Dada la gran cantidad de información que contienen estos hipercubos, mucha de ella redundante, es necesario el uso de técnicas de corrección espectral y estadísticas que ayuden al mejor aprovechamiento de la información que estas imágenes ofrecen.

### **4.2. Objetivos**

El objetivo general de este estudio fue el desarrollo de la metodología para la adquisición de imágenes hiperespectrales en semillas de uva, probando su potencial de manera preliminar, en la discriminación varietal y la predicción de la etapa de maduración. Los objetivos concretos planteados en este trabajo fueron:

- Optimización de la adquisición de imágenes hiperespectrales y tratamiento de los hipercubos para segmentación y obtención de la información bajo lenguaje de programación MATLAB.
- Aplicación de técnicas estadísticas multivariantes para la discriminación varietal y la predicción del estado de maduración.

### **4.3. Parte experimental**

Las muestras utilizadas fueron semillas procedentes de uvas tintas (Tempranillo y Syrah) y blancas (Zalema, cultivada en arena y barro), de la denominación de origen “Condado de Huelva”, tomadas en un intervalo de aproximadamente 70 días durante la maduración de la cosecha de 2011. De una muestra bruta de dos kilos de uva se tomaron al azar cien bayas. Y de ellas, se extrajeron las semillas con la ayuda de un cuchillo romo, se eliminaron los restos de pulpa y se mantuvieron en congelación hasta la adquisición de las imágenes.

Las imágenes hiperespectrales se adquirieron en el laboratorio del grupo *Food Refrigeration and Computerised Food Technology (FRCFT)* durante una



estancia en el *University College Dublin (UCD)*. El equipo de medida constaba de un espectrógrafo Inspector N17E de Specim® (Oulu, Finlandia), un dispositivo CCD Xeva 992 de Xenics® (Lovaina, Bélgica), una bandeja de muestra móvil MSA15R-N AMT-Linearways (SuperSlides & Bushes Corp, India), con dos lámparas de halogenuro de wolframio V-light (Lowel Light Inc, EE.UU.) como sistema de iluminación, y el software de adquisición de datos SpectralCube® de Specim® (Oulu, Finlandia).

El primer paso para la optimización de la metodología fue decidir la disposición más adecuada de las semillas en la bandeja de muestra para la adquisición de imágenes. Se probaron dos configuraciones, una de ellas con las semillas agrupadas en una cápsula de polipropileno, y otra colocando las semillas esparcidas directamente sobre la bandeja de muestra (Figura 32). Se observó que dentro de la cápsula de polipropileno, al estar unas sobre otras, las semillas producían sombras que se manifestaban en incongruencias en las intensidades de los espectros en los diferentes píxeles. Se decidió colocar las semillas esparcidas directamente en la bandeja como configuración seleccionada para la adquisición de imágenes ya que, además de evitar las sombras, facilitaba la segmentación, maximizaba la superficie de muestra en la imagen y mejoraba, por tanto, la representatividad.

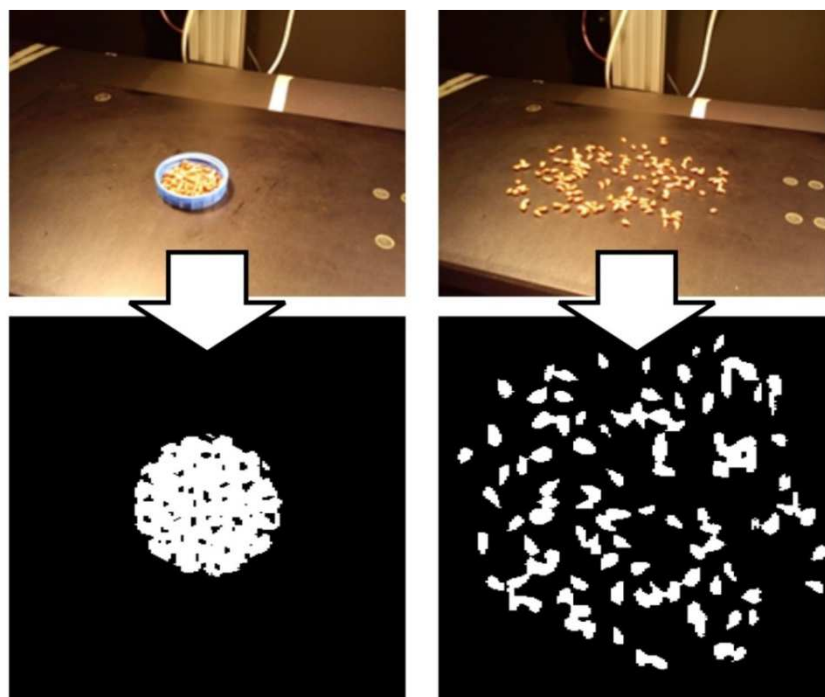


Figura 32. Distintas configuraciones para la adquisición de hipercubos y máscaras de segmentación resultantes.

En primer lugar, se calibraron las imágenes con un blanco de referencia, y se segmentaron mediante un método basado en umbralización a partir del resultado de operar matemáticamente con planos del hipercubo a diferentes longitudes de onda. Para ello, se seleccionó una longitud de onda en la cual la reflectividad fuera la misma para las semillas y el fondo (1127.9 nm) y otra donde la diferencia de reflectividades fuera máxima (1446.4 nm). La diferencia entre ambas imágenes resultó una tercera donde se maximizaba la diferencia de reflectividad de las dos zonas y donde se pudo por umbralización seleccionar fácilmente la región de interés. A la máscara de segmentación resultante se le aplicó, además, una erosión con una matriz  $1 \times 1$  para eliminar de la máscara todos los píxeles pertenecientes a los bordes de las semillas, donde la curvatura de la muestra y la influencia del fondo podría perturbar la medición (González Marcos et al., 2006). Se consideró el espectro de cada muestra como la media de los espectros de reflectancia de todos los píxeles de la región de interés. Se aplicaron también pretratamientos espectrales como la tipificación de la absorbancia (SNV), corrección del efecto multiplicativo de la dispersión (MSC) o la derivación para

---

corregir los espectros de posibles efectos provocados por fenómenos de dispersión de la luz (*scattering*) de origen inherente a la muestra o fenómenos instrumentales.

Como análisis preliminar se aplicó una regresión por mínimos cuadrados parciales (PLSR) considerando los espectros de las muestras (pretratados y sin pretratar) como variables independientes (matriz  $X$ ) y la fecha de muestreo como variable dependiente (matriz  $Y$ ). Con ello se pretende observar si los espectros contenían información sobre el avance de la madurez independientemente de su variedad. Los coeficientes de determinación y los errores, tanto en la calibración como en la validación cruzada, fueron prácticamente los mismos en los espectros pretratados que en los espectros sin pretratar, lo que indica que no existieron importantes inconsistencias en la respuesta instrumental, así como en las propiedades físicas de las muestras. La influencia de la fecha de muestreo en la información espectral sugiere que los cambios que las semillas sufren durante la maduración pueden determinarse a partir de estos espectros. Este hecho no se pudo comprobar por carecer de los correspondientes datos químicos para estas muestras.

La selección de las variables que contienen la mayor parte de la información puede ser igual o más eficiente que la consideración del espectro completo (Wold et al., 1996). Para ello, se aplicó un análisis de componentes principales (PCA) a los espectros de las muestras sin tratar para la selección de las bandas de mayor influencia. A partir de los tres primeros PC (que explicaban el 99% de la varianza) se seleccionaron las seis longitudes de onda con mayor peso: una longitud de onda del PC1 (928 nm), tres del PC2 (1148, 1325 y 1620 nm) y dos del PC3 (940 y 1656 nm). Los enlaces químicos absorben radiación a longitudes de onda específicas, por lo que se puede extraer esta información a partir del espectro de reflectancia. En concreto, las bandas seleccionadas resultaron pertenecer a terceros (928 y 940 nm), segundos (1148 nm) y primeros (1620 y 1652) sobretonos de tensión de enlaces C-H (Osborne et al., 1993). Parece razonable asumir que los cambios de humedad que sufren las semillas durante la maduración afectan a estos espectros, incluso más de lo que lo hacen los compuestos fenólicos de interés.

Al representar las puntuaciones de los dos primeros PC no es posible distinguir entre variedades. No obstante, se observa que las muestras aparecen en cierta manera alineadas para cada variedad según la fecha de muestreo. Al representar entonces el primer PC frente a la fecha de muestreo se observa para todas las variedades una alta dependencia, que resulta ser más acentuada en las etapas iniciales. La aparición de la variable tiempo en el primer PC sugiere que existen más diferencias en las semillas debido al estado de maduración que debido a las diferentes variedades. Asumiendo que el PC1 no contenía información varietal, se representaron los factores del PC2 frente a los del PC3 agrupando según variedades y se encontró que el PC3 diferenciaba claramente las semillas de uva tinta de las semillas de uva blanca. Dentro de las tintas, las semillas de Tempranillo tuvieron puntuaciones más altas que Syrah. Incluso para la variedad blanca Zalema, fue posible diferenciar las semillas por tipo de suelo a partir del PC2. Debe remarcarse que en el estudio de discriminación llevado a cabo utilizando el sistema DigiEye® (Capítulo 1 de esta Memoria), se pudieron diferenciar las variedades a partir de las características colorimétricas y morfológicas, pero no se diferenció la variedad blanca por tipo de suelo, lo que supone un avance en la metodología.

Para probar la eficacia de la reducción de variables, se realizó un nuevo PCA, pero usando como variables independientes las seis longitudes de onda seleccionadas anteriormente (928, 940, 1148, 1325, 1620 y 1656 nm) en lugar de las 240 totales. En este caso, utilizando solamente un 2.5% de la información espectral, se consiguieron diferenciar con éxito las variedades y el tipo de suelo con la misma eficacia que con el espectro completo. La representación tridimensional de los tres primeros PC, considerando todo el espectro y las seis longitudes de onda seleccionadas, permite ver claramente los resultados descritos.

Una de las ventajas de las técnicas de imagen es la obtención de información de cada punto de la muestra. Aprovechando esta ventaja, se aplicó un PCA directamente en los píxeles en lugar de en los espectros medios, y se obtuvieron nuevas imágenes donde se representaban las puntuaciones en escala de grises para cada PC (Geladi et al., 1989). La imagen del PC1 mostraba el mismo

---

significado que considerando el espectro completo, y se vio cómo la intensidad de los píxeles decrecía drásticamente en los primeros estadios y se estabilizaba al final. No se encontró información varietal en este PC. Al representar las imágenes pertenecientes a los PC 2 y 3 se observaron diferencias en las intensidades (puntuaciones del PCA) según la variedad de semilla de uva y el tipo de suelo.

Estos resultados están en concordancia con artículos previos, donde se demuestra cómo la concentración de flavan-3-oles y taninos evolucionan en las semillas de uva durante la maduración (De Freitas & Glories, 1999; Santos-Buelga et al., 1995), y cómo los factores agronómicos afectan a la concentración de estos compuestos (Ristic & Iland, 2005). Por ejemplo, se ha demostrado que las semillas de Syrah tienen menos niveles de taninos que las semillas de Tempranillo (Ribereau-Gayon, Glories, Maujean, and Dubourdieu, 2006).

Para verificar el potencial de las bandas seleccionadas mediante la reducción de variables, y dentro del análisis exploratorio, se aplicó un análisis discriminante (GDA) utilizando independientemente las muestras con todo el espectro y solo con las seis longitudes de onda seleccionadas. Se observa en los resultados de ambas matrices de clasificación un 100% de los casos correctamente clasificados utilizando todos los datos del espectro. Este porcentaje disminuye al utilizar solamente las seis longitudes de onda, pero es aún satisfactorio, ya que la clasificación es acertada para más de un 96% de los casos.

#### **4.4. Conclusiones**

Se ha establecido con éxito la metodología para la evaluación de las semillas de uva mediante imágenes hiperespectrales. El método permite el uso de semillas sin tratamiento de muestra adicional, y los espectros adquiridos no necesitan preprocesado para la obtención de información analítica útil.

- Las técnicas estadísticas multivariantes son capaces de predecir el estado de maduración de las semillas independientemente de la variedad, y permiten alcanzar coeficientes de determinación  $R^2$  superiores a 0.95.

- Los métodos estadísticos PCA y GDA son apropiados para la caracterización varietal de las semillas de uva a partir de sus características espectrales. Incluso para la variedad blanca Zalema se pueden encontrar diferencias debidas al tipo de suelo de cultivo a partir de su espectro NIR.
- Las seis bandas seleccionadas en la reducción de variables resultan tan efectivas como el espectro completo, de hecho, coinciden con bandas de relevante información química, según estudios anteriores.



## Grape seed characterization by NIR hyperspectral imaging

Francisco J. Rodríguez-Pulido<sup>a</sup>, Douglas F. Barbin<sup>b</sup>, Da-Wen Sun<sup>b,\*</sup>, Belén Gordillo<sup>a</sup>,  
M. Lourdes González-Miret<sup>a</sup>, Francisco J. Heredia<sup>a</sup>

<sup>a</sup> Food Colour & Quality Laboratory, Dept. Nutrition & Food Science, Facultad de Farmacia, Universidad de Sevilla, 41012 Sevilla, Spain

<sup>b</sup> Food Refrigeration and Computerised Food Technology (FRCFT), School of Biosystems Engineering, University College Dublin, National University of Ireland, Agriculture & Food Science Centre, Belfield, Dublin 4, Ireland

### ARTICLE INFO

#### Article history:

Received 19 June 2012

Accepted 18 September 2012

#### Keywords:

Hyperspectral imaging

*Vitis vinifera*

Grape seeds

Partial least squares regression

Principal components analysis

### ABSTRACT

Currently, the time of grape harvest is normally determined according to the sugar level in the pulp of the berry. Nonetheless, the stage of maturation in grape seeds should be taken into account more frequently to decide the appropriate harvest period. There are chemical and sensory analyses available to assess stage of maturation of grape seeds but they are destructive and time-consuming. Hyperspectral imaging is an alternative technology to characterize the grape seeds according to their chemical attributes, and the current work aimed to non-destructively characterize grape seeds in regard of the variety and stage of maturation. For this purpose, 56 samples of seeds from two red grape varieties (Tempranillo and Syrah) and one white variety (Zalema) in two kinds of soil were selected to assess their features based on the reflectance in the near-infrared (NIR) spectra by using prediction models (partial least squares regression) and multivariate analysis methods (principal component analysis and general discriminant analysis). In this study, a reliable methodology for predicting the stage of maturation was developed, and it was shown that it was possible to distinguish the variety of grape and even the type of soil from hyperspectral images of grape seeds.

© 2012 Elsevier B.V. All rights reserved.

### 1. Introduction

The composition of the grape seeds changes during maturation until grapes reach ripeness, affecting the sensory properties of wine. In addition to being an essential factor during the winemaking, grape seeds are useful after this process for other purposes such as cosmetics, fertilizers, and food supplements (Arvanitoyannis et al., 2006; Ruggieri et al., 2009; Selani et al., 2011). Chemical analyses are the most widely accepted reference methods for evaluating seeds. However, these methods frequently are destructive, requiring lengthy preparation procedures, and entail large samples. Sensory analysis is another approach to assess the condition of the seeds. However, this technique is time consuming and it is difficult to be carried out in an objective manner (Rousseau and Delteil, 2000). Conventional near infrared spectroscopy has been already applied to grape seeds to determine their chemical composition. Nevertheless, spectroscopy by transmission requires lengthy and laborious preparation procedures such as freeze-drying and grinding, and extraction processes are needed before

sample presentation to the analytical equipment (Wu et al., 2008; Ferrer-Gallego et al., 2010).

Computer vision systems (Sun, 2000; Brosnan and Sun, 2002, 2004; Sun and Brosnan, 2003; Du and Sun, 2004, 2006; Zheng et al., 2006a, 2006b) have the limitation of only acting on the surface of materials, thus only allowing identification of external features. However, phenolic compounds such as catechin and epicatechin are mainly concentrated within the outer layer of grape seeds (Thorngate and Singleton, 1994). Near-infrared (NIR) hyperspectral imaging is a powerful technique which has been used in a number of applications in agricultural products (Baye et al., 2006; Cubero et al., 2011; Shahin and Symons, 2011; Baiano et al., 2012; Lorente et al., 2012). By combining both spatial and spectral features, this technique provides an alternative, non-destructive technology for measuring constituents of biological materials. Moreover, this technique provides a way to distinguish among varieties in foodstuffs that would be very difficult by visual inspection (Lacar et al., 2001; Choudhary et al., 2009). In hyperspectral imaging, utilization of NIR measurements allows simultaneous determination of multiple constituents in a sample, since organic molecules have specific absorption patterns in the NIR region that can be used to characterize the chemical composition of the substance being examined (Williams and Norris, 2001).

Hyperspectral imaging provides a high spectral resolution allowing for models to utilize the spectral information in

\* Corresponding author. Tel.: +353 1 7167342; fax: +353 1 7167493.

E-mail address: [dawen.sun@ucd.ie](mailto:dawen.sun@ucd.ie) (D.-W. Sun).

URLs: <http://www.ucd.ie/refrig> (D.-W. Sun), <http://www.ucd.ie/sun> (D.-W. Sun).

combination with the image data. This approach is useful in applications where spectral information may not be sufficient, where similar constituents need to be separated. NIR hyperspectral imaging systems acquire spectral images, also known as hypercubes, which are three-dimensional data matrix where the first two axes ( $x$  and  $y$ ) of the matrix represent the spatial coordinates, while the third ( $\lambda$ ) axis depicts the spectral dimension. Hundreds of single channel black and white (grayscale) images, where each image represents a single band of spectral wavelength, are stacked on top of each other to produce a hypercube (Burger and Geladi, 2006). NIR imaging thus has the potential application for early detection of chemical changes and can be superior to visual based systems and spectroscopic methods.

Hyperspectral images (HSI) are large in size and spectral data are highly correlated, requiring the application of multivariate data analysis techniques for data exploration. As with NIR spectroscopy, chemometric techniques are applied to decompose the image dataset, pre-process and perform regression or classification analyses. Identifying most useful wavelengths is a good means of reducing the large amounts of data.

Several studies have illustrated the potential of using hyperspectral imaging based on the NIR range to develop a model able to predict and classify barley kernels according to germination stage (Munck and Møller, 2004; Engelbrecht et al., 2010). Williams et al. (2009) employed PLS discriminant analysis to classify maize kernels into hardness categories. Combination of the visual and NIR spectral ranges (400–1000 nm) were used to classify  $\alpha$ -amylase activity into two classes and to detect sprout damage to Canada Western Red Spring wheat (Xing et al., 2010).

In NIR spectral measurements, the most commonly used methods for spectral correction are multiplicative scatter correction (MSC), standard normal variate (SNV) and derivation (Geladi et al., 1985; Isaksson and Næs, 1988; Kaihara et al., 2002; Pizarro et al., 2004; Windig et al., 2008). However, there is still no standard procedure to decide which spectral pre-processing produce best results to a given dataset.

Partial least squares regression (PLSR) is a procedure used to relate a large number of independent variables (predictors) to the prediction of one or few response variables (observations). This technique is particularly effective in spectral analysis since it reduces a great number of highly correlated original descriptors (wavelengths) to a new variable space based on orthogonal factors called latent variables (Lin et al., 2012; Pojić and Mastilović, 2012). Results of the leave-one-out cross-validation (LOOCV) and test set approaches were compared to corroborate the prediction model results. Performance of the prediction models was evaluated using the root mean square error of calibration (RMSEC), the coefficient of determination in calibration and cross-validation ( $R^2_c$  and  $R^2_{cv}$ ), the root mean square error of cross-validation (RMSECV), the prediction error sum of squares (PRESS) and number of latent variables required (LV) (Fawcett, 2006).

A PCA was conducted to analyze the variance among the spectral information obtained from the samples. The loadings resulting from PCA could also be used to identify and investigate the influence of the wavelengths (variables). General discriminant analysis (GDA) is another commonly used technique for data classification and dimensionality reduction. This method maximizes the ratio between-class variance to the within-class variance in any particular dataset thereby guaranteeing maximal separability (González-Miret et al., 2006; Johnson, 2000).

The main objective of the present study was to investigate the potential of using NIR hyperspectral reflectance imaging technique as a fast and non-invasive method to characterize grape seeds according to varieties and stage of maturation. Specific objectives were to (1) establish a satisfactory approach to extract spectral data from hyperspectral images of grape seeds acquired in the NIR range

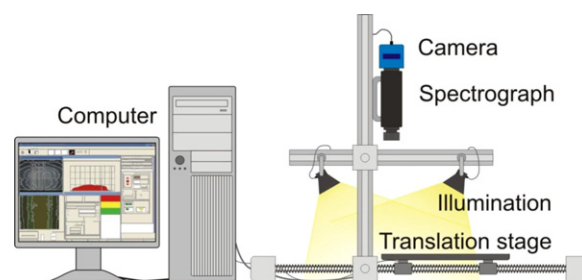


Fig. 1. Schematic diagram of the hyperspectral imaging system.

(900–1700 nm), (2) study whether spectral pre-processing methods can improve robustness of the prediction models, (3) identify the most significant wavelengths linked to the seed characteristics, (4) build robust PLSR models to relate spectral information and stage of maturation using selected wavelengths.

## 2. Materials and methods

### 2.1. Sample preparation

The vineyards sampled are included under the “Condado de Huelva” Designation of Origin, in Southwestern Spain, harvested in 2011. Fields in Condado de Huelva are flat or weakly undulated, and their medium fertility and neutral or slightly basic characteristics make of them an ideal land for growing grapes. Two red varieties (Tempranillo and Syrah) and one autochthonous white variety (Zalema) cultivated in two kinds of soil (sand and clay) were used. The number of sample points was from 12 to 17 depending on the different harvesting times of each variety (a total of 56 samples). Sandy soils contain about 75% of fine sand, and they frequently need surface work because of their tendency to form shells that avoid aeration. Their permeability is very good and they give high yields. Clay soils include Mediterranean red soils, and with heavy texture and low permeability, clay soil is less favorable to the vine than sandy soil. According to availability of each variety in vineyards, samples were taken twice a week from late June until postharvest early September. Sampling was carried out at daybreak by taking a pair of berries from alternate grapevines, from four rows of vines, and from both sides of each row up to reach 2 kg of berries. In this process, the berries were taken with pedicel intact to slow down the berry oxidation as long as possible. Once in the laboratory, 100 berries were randomly taken and seeds removed, left to dry at room temperature for 2 h, and frozen at  $-20^{\circ}\text{C}$  until acquisition of hyperspectral images.

### 2.2. Acquisition and processing of hyperspectral images

NIR spectral images were acquired in the reflectance mode using a pushbroom hyperspectral imaging system (Fig. 1). The system comprises a spectrograph (ImSpector, N17E, Spectral Imaging Ltd, Finland), a charged-couple device (CCD) camera with C-mount lens (Xeva 992, Xenics Infrared Solutions, Belgium), a translation stage (MSA15R-N, AMT-Linearways, SuperSlides&Bushes Corp., India) with two tungsten-halogen lamps (V-light, Lowell Light Inc, USA) providing illumination, data acquisition software (SpectralCube, Spectral Imaging Ltd., Finland), and a computer. The conveying translation stage was driven by a stepping motor (GPL-DZTSA-1000-X, Zolix Instrument Co, China) with a user-defined speed of  $2.7\text{ cm s}^{-1}$  synchronized with the image acquisition by the data acquisition software. The horizontal axis of the spatial dimension of the images ( $x$ ) was fixed, while the sample was scanned underneath the field of view (FOV) of the camera to determine the size of the vertical axis of the spatial dimension ( $y$ ) of the spectral images.



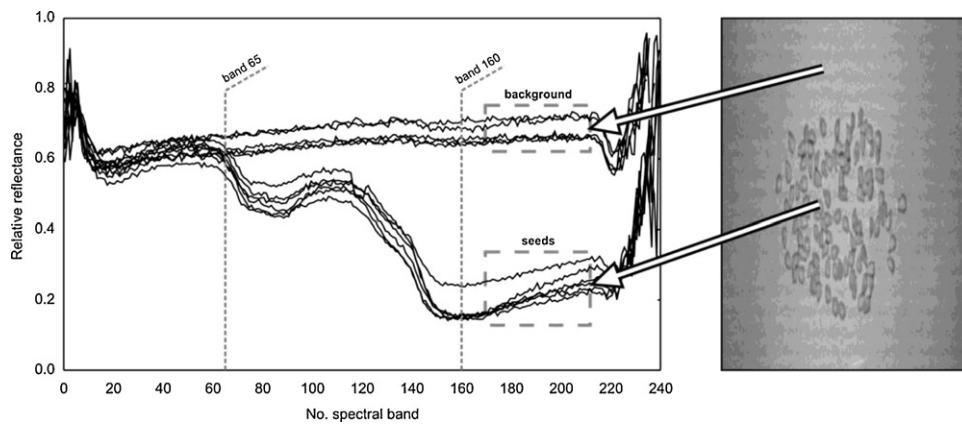


Fig. 2. Spectra of some pixels from background and seeds and selected bands to create the segmentation mask.

The conveyor speed was adjusted to fit the predetermined camera exposure time to avoid distortion on image size, providing identical spatial resolution of the vertical and horizontal axes. The spectral range ( $\lambda$ ) recorded was 897–1752 nm at an increment of 3.34 nm, producing hyperspectral images at 256 wavelength channels. The complete image acquisition process was controlled by the SpectralCube software (Spectral Imaging Ltd., Finland). The region of interest (ROI) was selected with MATLAB R2011b (The Mathworks, 2009) and all steps described for spectral analysis were carried out using multivariate analysis software (Unscrambler version 9.7, CAMO, 2007).

The seeds of each sample were presented to the system spread on the translation stage and conveyed to the FOV of the camera to be scanned line by line. By examining the acquired hyperspectral images, it was observed that the first five and the last eleven bands of the image had a high level of noise, thus not being useful for spectral data extraction. Therefore, images were cropped to the spectral range of 914–1715 nm with a total of 240 bands.

A 'white reference' image ( $W$ , 100% reflectance) was acquired from a white reference ceramic tile, and a 'dark reference' image ( $B$ , 0% reflectance) was obtained with the light source off and the camera lens completely covered with its opaque cap. The white and dark 'reference' hyperspectral images were used to correct the raw images ( $R_0$ ) to obtain a relative reflectance image ( $R$ ) according to the following equation:

$$R = \frac{R_0 - B}{W - B} \quad (1)$$

Once the images were cropped and corrected as described above, spectra belonging to background and seeds were overlaid. One band was selected with low spectral reflectance difference between background and seeds and another band with maximum

spectral reflectance difference between background and seeds. These bands were 65 and 160 (1127.9 and 1446.4 nm, respectively) (Fig. 2). By subtracting these two images it is possible to obtain a new image where this difference appears in grayscale. A simple thresholding procedure identified for pixels with the difference higher than 0.2, which were reassigned with a value of one to form a binary mask (Fig. 3). Moreover, the shape of the seeds was eroded with a  $1 \times 1$  matrix in order to avoid considering pixels in the edges with low-intensity spectrum (González Marcos et al., 2006). For each sample, the region considered by the binary mask was used for the calculation of the mean spectrum as the average of spectra of all pixels within this region. An average spectrum was extracted from each sample by using the segmentation criterion. Thus, a total of 56 mean reflectance spectra were extracted from grape seed samples from different varieties.

According to Wold et al. (1996), selecting optimum wavelengths that carry most of the information may be equally or more efficient than using full spectra. The reduced number of wavelengths is enough to characterize most of the classification tasks (Vila et al., 2005). In this study, PCA loadings were used for identification of optimal wavelengths that have a high influence in each PC (Lawrence et al., 2004). Loadings resulting from PCA of the spectral data of all samples represent the regression coefficients for each wavelength at the respective principal component and indicate the most dominant wavelengths. New general discriminant analysis (GDA) was carried out using only a few selected wavelengths, and the results were compared with the classification obtained by using the whole spectra.

Pre-processing techniques were implemented to analyze the impact of systematic noise in the spectral data acquired from the grape seeds, while retaining useful information for sample characterization. This step aimed to establish a methodology that would

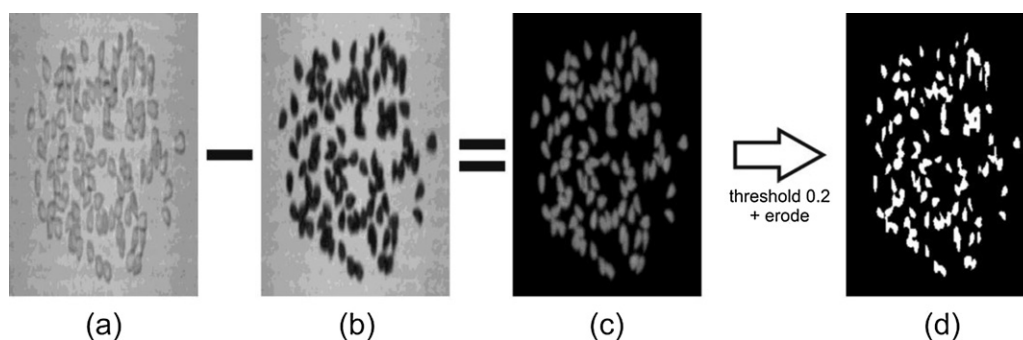


Fig. 3. Segmentation steps: (a) image of relative reflectance at 1127.9, (b) and 1446.4 nm, (c) image resulting from the subtraction of the two images, and (d) the resulting binary mask after thresholding segmentation and erosion.

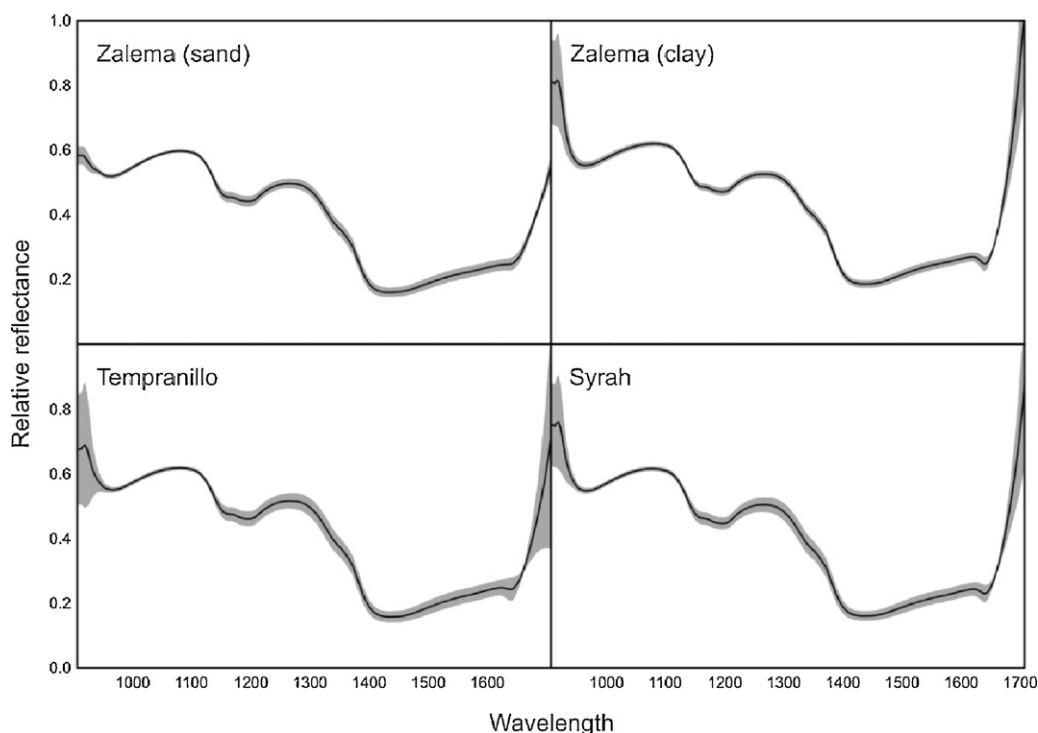


Fig. 4. Average spectra for the respective samples of each variety. Shaded area represents the standard deviation in each wavelength.

allow a systematic approach for further studies. PLSR was applied to the raw spectral datasets (240 bands) with full cross-validation (leave-one-out) for the original raw spectral data and for pre-processed spectra to predict the stage of maturation. In addition, 32 samples (60% of the total number of samples) were randomly selected as a training set (calibration set) and 22 samples (40% of the total number of samples) were used as a validation set (prediction set). A *t*-test was carried out comparing both sets to ensure that it covered appropriately and consistently the whole range of values for each response, since the same set was used for all the pre-processing methods.

### 3. Results and discussion

Fig. 4 shows the mean and standard deviation spectrum for each class of grape seeds. It can be seen that different categories have different reflectance intensities along some wavelength regions, although with the same pattern. Usually the high reflectance around 950 nm can be attributed to the high water content of biological materials (Murray and Williams, 1987; Osborne et al., 1993). For each sample, the average spectrum obtained from pixels of HSI in seeds was used for further statistical treatment.

The results obtained for raw data and each pre-processing treatment selected are summarized in Table 1. In all cases, the optimal

**Table 1**  
Prediction results for raw spectra and pre-processed spectra from samples of grape seeds.

	# LV	$R^2_C$	RMSEC	$R^2_{CV}$	RMSECV
Raw	5	0.95	4.06	0.94	4.88
MSC	5	0.96	3.82	0.94	4.58
SNV	5	0.95	4.18	0.94	4.78
1st derivative	5	0.95	4.30	0.93	5.13
2nd derivative	5	0.95	4.21	0.91	5.54
Test set	5	0.95	4.31	0.93	4.09

number of LV for establishing the calibration model was five. The coefficient of determination in calibration and cross-validation was high and almost the same in all cases. Moreover, there was no difference between predictions in raw spectra and predictions in pre-processed spectra. This suggests that there were no important inconsistencies in instrument response as well as in physical characteristics in samples.

The different pre-processing methods facilitated to investigate the performance of the PLSR models. The difference between  $R^2_C$  and  $R^2_{CV}$  was small and could indicate the lack of overfitting in the model. The coefficient of prediction values obtained for the full cross-validation and test set validation models were comparable in terms of prediction performance. Results presented in Table 1 show that regression models based on pre-processed spectra performed similarly to the models for the raw spectra. Derivative

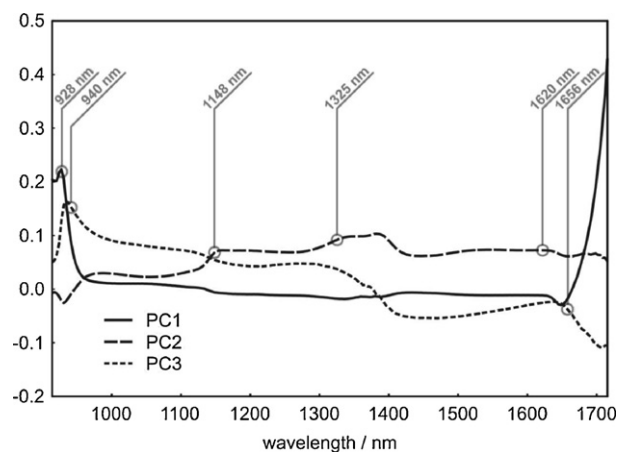


Fig. 5. Loadings of the first three principal components showing the selected wavelengths.

transformation provided equivalent results compared to other pre-processing methods, with comparable  $R^2_{CV}$  and RMSECV for the prediction models.

Nevertheless, none of the aforementioned pre-treatment methods resulted in significant improvement of predictive ability in comparison to the raw spectral data, thus the negative impact of the scatter effects on the regression quality was not confirmed. Therefore, it can be assumed from these results that the applied pre-treatments were not effective for this particular dataset. Given that the predictive ability of the models was similar to that obtained with the original data, it is feasible to assume that the raw spectra extracted using the proposed segmentation approach can be used as representative sample for further analyses. The suitability of PLSR for predicting properties from NIR hyperspectral images of foodstuffs agrees with other similar studies (Menesatti et al., 2009; Barbin et al., 2012).

The loadings of the first three principal components were used for wavelength selection because these three principal components were responsible for 99% of the variance in the spectral data. The wavelengths corresponding to higher module values (peaks and valleys) at these particular principal components were selected as candidates for optimum wavelengths (Fig. 5). Six optimum wavelengths (928, 940, 1148, 1325, 1620 and 1656 nm) were thus identified for discrimination purposes.

PC1 explains 95.99% of the total variance in the samples but only one wavelength (928 nm) was selected from this component, while three wavelengths were selected from PC2 (1148, 1325 and 1620 nm), and two from PC3 (940 and 1656 nm). Due to the fact that chemical bonds absorb light energy at specific wavelengths, some compositional information can be determined from the reflectance spectra. The spectral information is repeated through the successive overtones and combination regions. Five of the six optimum wavelengths selected by PCA loadings are due to C–H stretching third (928 and 940 nm), second (1148 nm) and first overtones (1620 and 1652 nm), respectively (Osborne et al., 1993). It could be reasonable to assume that change in water content and other components in the samples were detected influencing the NIR spectra and causing variation during the maturation. However, the NIR spectrum contains information from all the chemical constituents of the sample and direct interpretation of the spectral reflectance values is difficult for complex materials such as intact crop seeds (Williams and Norris, 2001). Furthermore, chemical analyses should be performed to confirm the variation in chemical composition among the samples that were observed in the NIR spectra.

Fig. 6 shows the scores for PC1 and PC2. In this graph, some differences were observed, but it is not easy to distinguish among varieties. Nevertheless, a trend in time can be observed along  $x$ -axis (PC1). In this sense, the stage of maturation could be explained by the first principal component. In the plot of PC1 vs time (Fig. 7), the dependency can be expressed by its slope. A high slope (in absolute values) means that PC1 is influenced by the time in a more extensively manner. This way, the changes in this principal component are greater in the early stages. Except for Zalema (sand), there is a high dependency between PC1 scores and sampling date. In order to distinguish among varieties, the scores for PC2 and PC3 were also plotted. In Fig. 8a, seeds from red grapes (Tempranillo and Syrah) have positive scores in PC3 while seeds from white variety (Zalema) have negative scores. The third principal component also divides the red varieties in two clusters with Tempranillo having higher scores than Syrah. Regarding the variety of Zalema, PC2 splits the grape seeds into two clusters according to the kind of soil (sand and clay). As shown in Fig. 8b, the scores from PCA using the six selected wavelengths divides the red and white varieties similarly using the whole spectrum. However, the division between the two red varieties is not as clear as by using full spectra. A clearer classification

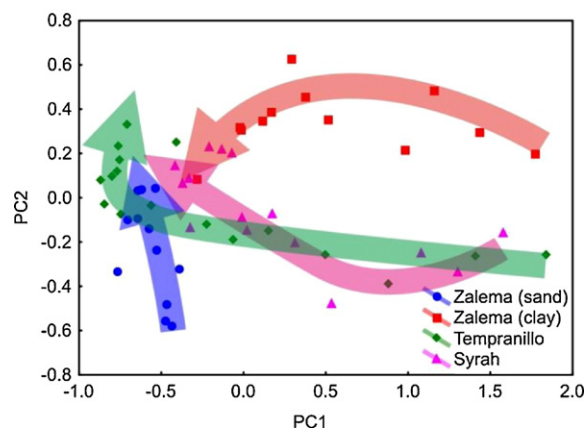


Fig. 6. Scatterplot of scores for PC1 and PC2. The direction of the arrows indicates the trend in time in each variety.

can be observed in Fig. 9. It shows a three-dimensional scatterplot containing the three first principal components. The principal component analysis successfully separated the four types of samples used in this study. Although there was a lack of accuracy in some samples, this hyperspectral technique could distinguish two treatments such as kind of soil as it had been previously demonstrated (Karimi et al., 2012).

Since hyperspectral imaging contains the spatial distribution of reflectance, principal component analysis was also applied in images instead of the average spectra in each sample. In this case, scores appear as intensity in grey-scale in each image. The first principal components have the same meaning as that by considering the average spectrum (Fig. 10). PC1 describes change in time (seeds appear clearly brighter in the first stages and become darker in the last ones). Once more, this change was very weak for this component in Zalema (sand). PC2 also weakly described change in time. However, PC2 together with PC3 is useful for classification among varieties.

The spectral differences among varieties are in accordance with previous reports. The more important compounds in seed coat are monomers of flavan-3-ols and tannins. Their concentrations change during the maturation stages (Santos-Buelga et al., 1994; de Freitas and Glories, 1999). In addition, seed varieties and agronomic factors affect the concentration of these compounds (Ristic and Iland, 2004). For instance, Syrah fruit have a reduced level of seeds tannins in comparison to Tempranillo fruit (Ribereau-Gayon et al., 2000).

To verify the potential of selected optimum wavelengths for grape seed discrimination, GDA was conducted on the reflectance spectral data using the full spectral range (240 wavelengths) and only the optimum wavelengths selected (6 wavelengths).

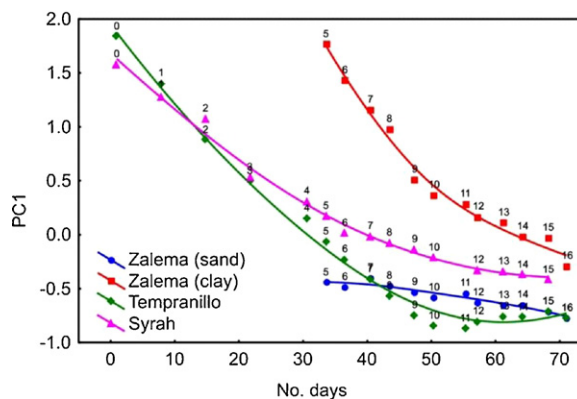


Fig. 7. Dependency of PC1 with date.

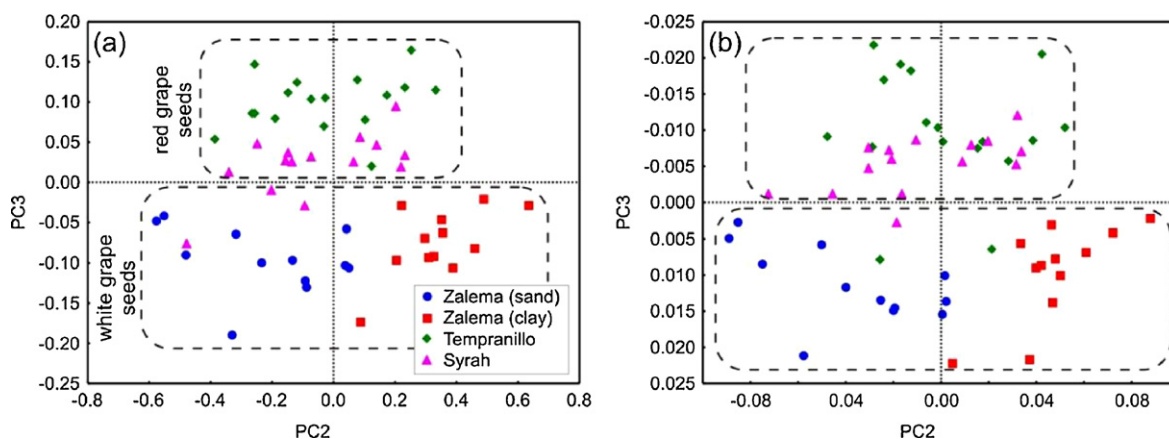


Fig. 8. Scatterplots of scores for PC2 and PC3 (a) using full spectra (240 bands) and (b) using the selected wavelengths (6 bands).

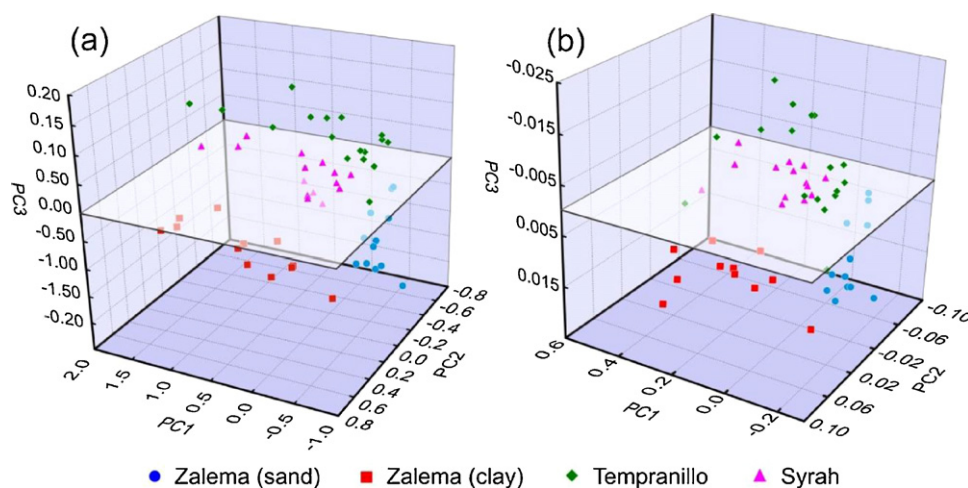


Fig. 9. Three-dimensional scatterplots of scores for PC1, PC2, and PC3 (a) using full spectra and (b) using the selected wavelengths (6 bands).

Tables 2 and 3 show the classification matrix for the GDA of grapes among varieties. Using the full spectra, it was possible to classify 100% of the samples according to their variety. The result using only six selected wavelengths is lower, although still satisfactory, since it provided an accuracy of more than 96%. It is clear that optimum wavelengths have a great discrimination power for

distinguishing among grape seed varieties. Acquiring images at those particular wavelengths would reduce image acquisition and processing time and could be useful for establishing a multispectral system for further studies. Due to the similar appearance in red grapes, the unique spectral profile of each variety could be a practical tool to assess the authenticity among varieties (Woodcock et al.,

Table 2  
Classification matrix after GDA using full spectra (240 bands).

	Percent	Zalema (sand)	Zalema (clay)	Tempranillo	Syrah
Zalema (sand)	100	12	0	0	0
Zalema (clay)	100	0	12	0	0
Tempranillo	100	0	0	17	0
Syrah	100	0	0	0	15
Total	100	12	12	17	15

Table 3  
Classification matrix after GDA using only the selected wavelengths (6 bands).

	Percent	Zalema (sand)	Zalema (clay)	Tempranillo	Syrah
Zalema (sand)	100.0	12	0	0	0
Zalema (clay)	91.7	0	11	0	1
Tempranillo	94.1	0	0	16	1
Syrah	100.0	0	0	0	15
Total	96.4	12	11	16	17

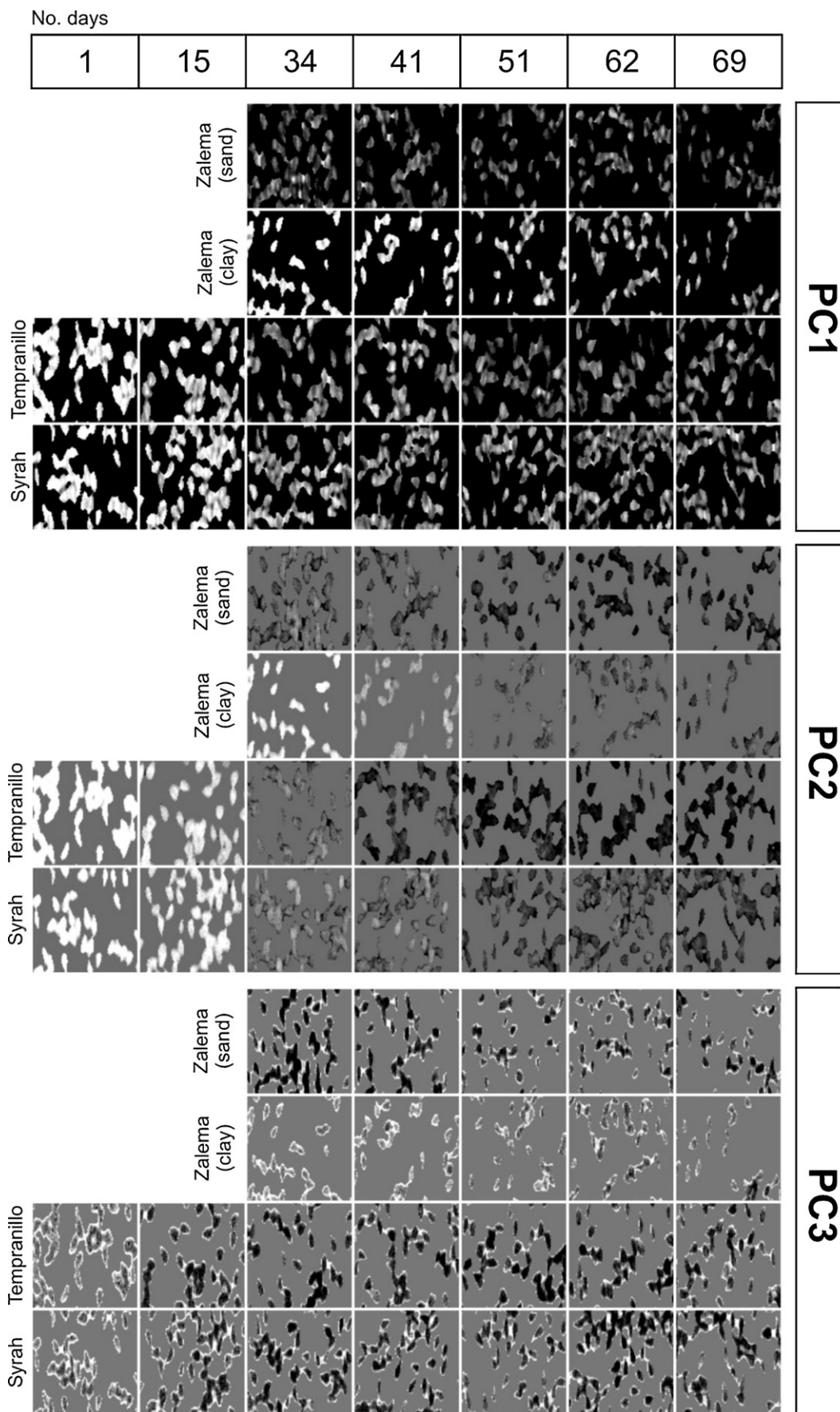


Fig. 10. First three score images obtained from PCA for the hyperspectral images of grape seeds. Horizontal scale shows evolution along maturation.

2008). This could be useful in routine inspections by Designations of Origin to wineries.

#### 4. Conclusions

The methodology for hyperspectral images acquisition and pre-processing in grape seeds was established. Likewise, the sample presentation as intact seeds and the segmentation criterion chosen provided a suitable way for extracting the mean spectrum of each set of seeds in one sample. The PLSR model applied was able to predict the stage of maturation of a sample based on spectral features as the predictor variables with the coefficients of determination being higher than 0.95. Moreover, the proposed image acquisition method eliminated the need for pre-processing of images for correction on scattered pixels.

Both PCA and GDA methods were suitable for grape seed characterization according to their varieties. Within the same variety (Zalema), these methods could distinguish between the two kinds of soil where vines were cultivated. The bands containing the most relevant chemical information identified in the PCA loadings are in accordance to previous results reported in the literature.

This work has shown the capability of hyperspectral imaging to predict the variety and stage of maturation of the samples analyzed without additional analytical analyses, since these features are influenced by changes in chemical composition. However, reference chemical tests should be further necessary, since image analysis is an indirect method. On the other hand, hyperspectral imaging arises as a useful alternative since it can overcome some disadvantages of the traditional analytical methods such as lack of versatility and monotonous sample preparation. Hyperspectral imaging allows the evaluation of small samples with interactive selection of regions of interest, something that is not possible with conventional spectroscopy. In combination with the spectral features obtained for the samples, it has the potential to become a useful tool not only in winery industry but also in other postharvest applications where assessing grape seeds attributes are necessary.

#### Acknowledgements

This work was supported by the projects AGL2011-30254-C02 (Ministerio de Economía y Competitividad, Gobierno de España), P10-AGR6331 (Consejería de Economía, Innovación, Ciencia y Empresa, Junta de Andalucía), and the concession of the fellowship (BES-2009-025429).

#### References

- Arvanitoyannis, I.S., Ladas, D., Mavromatis, A., 2006. Potential uses and applications of treated wine waste: a review. *International Journal of Food Science and Technology* 41, 475–487.
- Baiano, A., Terracone, C., Peri, G., Romaniello, R., 2012. Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes. *Computers and Electronics in Agriculture* 87, 142–151.
- Barbin, D.F., ElMasry, G., Sun, D.W., Allen, P., 2012. Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging. *Analytica Chimica Acta* 719, 30–42.
- Baye, T.M., Pearson, T.C., Settles, A.M., 2006. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *Journal of Cereal Science* 43, 236–243.
- Brosnan, T., Sun, D.-W., 2002. Inspection and grading of agricultural and food products by computer vision systems - a review. *Computers and Electronics in Agriculture* 36 (2–3), 193–213.
- Brosnan, T., Sun, D.-W., 2004. Improving quality inspection of food products by computer vision - a review. *Journal of Food Engineering* 61 (1), 3–16.
- Burger, J., Geladi, P., 2006. Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst* 131, 1152–1160.
- CAMO, 2007. Unscrambler. CAMO, Trondheim, Norway.
- Choudhary, R., Mahesh, S., Paliwal, J., Jayas, D.S., 2009. Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosystems Engineering* 102, 115–127.
- Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., Blasco, J., 2011. Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food and Bioprocess Technology* 4 (4), 487–504.
- de Freitas, V.A.P., Glories, Y., 1999. Concentration and compositional changes of pro-cyanidins in grape seeds and skins of white *Vitis vinifera* varieties. *Journal of the Science of Food and Agriculture* 79, 1601–1606.
- Du, C.J., Sun, D.-W., 2004. Recent developments in the applications of image processing techniques for food quality evaluation. *Trends in Food Science & Technology* 15 (5), 230–249.
- Du, C.J., Sun, D.-W., 2006. Learning techniques used in computer vision for food quality evaluation: a review. *Journal of Food Engineering* 72 (1), 39–55.
- Engelbrecht, P., Manley, M., Williams, P.J., Toit, G.D., Geladi, P., 2010. Pre-germination detected in whole cereal grains using near infrared hyperspectral imaging. In: *Proceedings of the CST SA – ICC International Grains Symposium Quality and Safety of Grain Crops and Foods*, pp. 123–127.
- Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27, 861–874.
- Ferrer-Gallego, R., Hernández-Hierro, J.M., Rivas-Gonzalo, J.C., Escribano-Bailón, M.T., 2010. Feasibility study on the use of near infrared spectroscopy to determine flavanols in grape seeds. *Talanta* 82, 1778–1783.
- Geladi, P., MacDougall, D., Martens, H., 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy* 39, 491–500.
- González Marcos, A., Martínez de Pisón Ascacibar, F.J., Pernía Espinoza, A.V., Alba Elías, F., Castejón Limas, M., Ordieres Meré, J., Vergara González, E., 2006. Segmentación. In: *Grupo de Investigación EDMANS (Eds.), Técnicas y algoritmos básicos de visión artificial*. Universidad de La Rioja, Servicio de Publicaciones, Logroño, pp. 55–72.
- González-Miret, M.L., Escudero-Gilete, M.L., Heredia, F.J., 2006. The establishment of critical control points at the washing and air chilling stages in poultry meat production using multivariate statistics. *Food Control* 17, 935–941.
- Isaksson, T., Næs, T., 1988. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Applied Spectroscopy* 42, 1273–1284.
- Johnson, D.E., 2000. *Métodos multivariados aplicados al análisis de datos*. International Thomson Editores, S.A., Madrid.
- Kaijara, M., Takahashi, T., Akazawa, T., Sato, T., Takahashi, S., 2002. Application of near infrared spectroscopy to rapid analysis of coals. *Spectroscopy Letters* 35, 369–376.
- Karimi, Y., Maftoonzad, N., Ramaswamy, H., Prasher, S., Marcotte, M., 2012. Application of hyperspectral technique for color classification of avocados subjected to different treatments. *Food and Bioprocess Technology* 5, 252–264.
- Lacar, F.M., Lewis, M.M., Grierson, I.T., 2001. Use of hyperspectral reflectance for discrimination between grape varieties. In: *IEEE 2001, International Geoscience and Remote Sensing Symposium, 2001 (IGARSS'01)*, vol. 6, pp. 2878–2880.
- Lawrence, K.C., Windham, W.R., Park, B., Smith, D.P., Poole, G.H., 2004. Comparison between visible/NIR spectroscopy and hyperspectral imaging for detecting surface contaminants on poultry carcasses. In: *Bennedsen, B.S., Chen, Y.R., Meyer, G.E., Senecal, A.G., Tu, S.I. (Eds.), Monitoring Food Safety, Agriculture, and Plant Health*. Proceeding of SPIE, Providence, RI, pp. 35–42.
- Lin, P., Chen, Y., He, Y., 2012. 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.
- Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O.L., Blasco, J., 2012. Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology* 5 (4), 1121–1142.
- Menesatti, P., Zanella, A., D'Andrea, S., Costa, C., Paglia, G., Pallottino, F., 2009. Super-vised multivariate analysis of hyper-spectral NIR images to evaluate the starch index of apples. *Food and Bioprocess Technology* 2, 308–314.
- Munck, L., Möller, B., 2004. A new germinative classification model of barley for prediction of malt quality amplified by a near infrared transmission spectroscopy calibration for vigour on line both implemented by multivariate data analysis. *Journal of the Institute of Brewing* 110, 3–17.
- Murray, I., Williams, P.C., 1987. Chemical principles of near-infrared technology. In: *Williams, P., Norris, K. (Eds.), Near-Infrared Technology in the Agricultural and Food Industries*. American Society of Cereal Chemists, St. Paul, MN, pp. 17–34.
- Osborne, B.G., Fearn, T., Hindle, P.T., 1993. *Practical NIR spectroscopy with applications in food and beverage analysis*, 2nd ed. United Kingdom, Longman Group.
- Pizarro, C., Esteban-Díez, I., Nistal, A.J., González-Sáiz, J.-M., 2004. Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta* 509, 217–227.
- Pojić, M.M., Mastilović, J.S., 2012. Near infrared spectroscopy –advanced analytical tool in wheat breeding, trade, and processing. *Food and Bioprocess Technology*, <http://dx.doi.org/10.1007/s11947-012-0917-3>.
- Ribereau-Gayon, P., Glories, Y., Maujean, A., Dubouard, D., 2000. Phenolic compounds. In: *The Chemistry of Wine Stabilisation and Treatments*. John Wiley & Sons, Paris.
- Ristic, R., Iland, P.G., 2004. Relationships between seed and berry development of *Vitis vinifera* L. cv Shiraz: developmental changes in seed morphology and phenolic composition. *Australian Journal of Grape and Wine Research* 11, 43–58.
- Rousseau, J., Delteil, D., 2000. Présentation d'une méthode d'analyse sensorielle des raisins. Principe, méthode et grille d'interprétation. *Revue Française D'oenologie* 183, 10–13.
- Ruggieri, L., Cadena, E., Martínez-Blanco, J., Gasol, C.M., Rieradevall, J., Gabarrell, X., Gea, T., Sort, X., Sánchez, A., 2009. Recovery of organic wastes in the Spanish wine

- industry. Technical, economic and environmental analyses of the composting process. *Journal of Cleaner Production* 17, 830–838.
- Santos-Buelga, C., Francia-Aricha, E.M., Escribano-Bailón, M.T., 1994. Comparative flavan-3-ol composition of seeds from different grape varieties. *Food Chemistry* 53, 197–201.
- Selani, M.M., Contreras-Castillo, C.J., Shirahigue, L.D., Gallo, C.R., Plata-Oviedo, M., Montes-Villanueva, N.D., 2011. Wine industry residues extracts as natural antioxidants in raw and cooked chicken meat during frozen storage. *Meat Science* 88, 397–403.
- Shahin, M.A., Symons, S.J., 2011. Detection of *Fusarium* damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis. *Computers and Electronics in Agriculture* 75, 107–112.
- Sun, D.W., 2000. Inspecting pizza topping percentage and distribution by a computer vision method. *Journal of Food Engineering* 44 (4), 245–249.
- Sun, D.-W., Brosnan, T., 2003. Pizza quality evaluation using computer vision – part 1 – Pizza base and sauce spread. *Journal of Food Engineering* 57 (1), 81–89.
- The Mathworks, 2009. MATLAB. The MathWorks Inc., Natick, MA.
- Thorngate, J.H., Singleton, V.L., 1994. Localization of procyanidins on grape seeds. *American Journal of Enology and Viticulture* 45, 259–262.
- Vila, J., Calpe, J., Pla, F., Gómez, L., Connell, J., Marchant, J., Calleja, J., Mulqueen, M., Muñoz, J., Klaren, A., 2005. SmartSpectra: applying multispectral imaging to industrial environments. *Real-Time Imaging* 11, 85–98.
- Williams, P., Geladi, P., Fox, G., Manley, M., 2009. Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta* 653, 121–130.
- Williams, P., Norris, K., 2001. *Near-Infrared Technology: In the Agricultural and Food Industries*, 2nd ed. American Association of Cereal Chemists, St Paul, MN.
- Windig, W., Shaver, J., Bro, R., 2008. Loopy MSC: a simple way to improve multiplicative scatter correction. *Applied Spectroscopy* 62, 1153–1159.
- Wold, J.P., Jakobsen, T., Krane, L., 1996. Atlantic salmon average fat content estimated by near-infrared transmittance spectroscopy. *Journal of Food Science* 61, 74–77.
- Woodcock, T., Fagan, C., O'Donnell, C.P., Downey, G., 2008. Application of near and mid-infrared spectroscopy to determine cheese quality and authenticity. *Food and Bioprocess Technology* 1, 117–129.
- Wu, D., He, Y., Feng, S., Sun, D.-W., 2008. Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. *Journal of Food Engineering* 84 (1), 124–131.
- Xing, J., Symons, S., Shahin, M., Hatcher, D., 2010. Detection of sprout damage in Canada Western Red Spring wheat with multiple wavebands using visible/near-infrared hyperspectral imaging. *Biosystems Engineering* 106, 188–194.
- Zheng, C.X., Sun, D.-W., Zheng, L.Y., 2006a. Recent developments and applications of image features for food quality evaluation and inspection—a review. *Trends in Food Science & Technology* 17 (12), 642–655.
- Zheng, C.X., Sun, D.-W., Zheng, L.Y., 2006b. Recent applications of image texture for evaluation of food qualities—a review. *Trends in Food Science & Technology* 17 (3), 113–128.





---

## CAPÍTULO 5

### ESTUDIO DE LA RELACIÓN ENTRE LA COMPOSICIÓN QUÍMICA Y EL PERFIL HIPERESPECTRAL EN LAS SEMILLAS DE UVA

Francisco J. Rodríguez-Pulido, José Miguel Hernández-Hierro, Julio Nogales-Bueno, Belén Gordillo, M. Lourdes González-Miret, Francisco J. Heredia. *A novel method for evaluating flavanols in grape seeds by near infrared hyperspectral imaging*. Talanta, ISSN: 0039-9140. (en revisión)

#### 5.1. Antecedentes

A pesar de constituir menos del 5% del peso total de la baya, las semillas de uva contribuyen extensamente a las propiedades del vino. Esta aportación se debe a los compuestos fenólicos que se desarrollan en ellas a lo largo de toda la maduración y que pueden cederse durante la vinificación. De ellos, los más representativos son los flavanoles, compuestos por monómeros de flavan-3-oles (catequina, epicatequina y epicatequin-galato) y procianidinas (Ristic & Iland, 2005).

Otro aspecto de relevancia, que está ganando interés en los últimos años, es la extractabilidad de estos compuestos, pues la importancia no está en la cantidad de flavanoles contenidos en las semillas sino en qué cantidad pasan al mosto durante el proceso de encubado. Los cambios en la extractabilidad radican en la modificación de la interacción entre los taninos y las paredes celulares, en los tegumentos externo e interno de las semillas. Estos cambios afectan tanto a las propiedades gustativas como a la apariencia de las semillas. Por tanto, el análisis sensorial es el método más común para evaluar las semillas por los viticultores, a pesar de la dificultad de llevar a cabo este análisis de una manera precisa y objetiva (Rousseau & Delteil, 2000).

Por otro lado, para evaluar la composición de las semillas objetivamente, el único modo existente hasta ahora es el análisis químico, que implica el uso de reactivos y un tiempo más o menos prolongado, además de destruir la muestra para posteriores análisis. En este sentido, el reemplazo de los análisis químicos

convencionales por técnicas espectroscópicas resulta una alternativa rápida, precisa y no destructiva para conocer la composición química (Hernández-Hierro et al., 2012; Nicolai et al., 2007). Además, han sido implementadas satisfactoriamente dentro de las técnicas de imagen (Barbin et al., 2013; Sun, 2004). Las imágenes hiperespectrales en NIR se han utilizado satisfactoriamente para la predicción de características químicas en muestras pequeñas y discretas, como son granos de cebada (Munck & Møller, 2004), maíz (Williams et al., 2009) y trigo (Xing et al., 2009; 2010).

### **5.2. Objetivos**

El objetivo general de este trabajo ha sido aplicar la metodología de análisis de imágenes hiperespectrales para la predicción del contenido en flavanoles de las semillas de uva a partir de sus espectros de reflectancia NIR. Los objetivos concretos planteados fueron:

- Determinación del contenido en flavanoles y su extractabilidad en semillas de uva. Para ello se seleccionaron semillas de diferentes variedades y en diferentes estadios de maduración.
- Medida del espectro infrarrojo de las semillas de uvas mediante la metodología de análisis de imágenes hiperespectrales, propuesta en el Capítulo 4 de esta Memoria.
- Aplicación de técnicas estadísticas multivariantes para predecir la composición química de las semillas de uva a partir de los correspondientes espectros de reflectancia en el intervalo infrarrojo cercano.

### **5.3. Parte experimental**

Para este estudio, se seleccionaron semillas de uva de las variedades tinta Tempranillo y blanca Zalema, de la Denominación de Origen Condado de Huelva, en la cosecha de 2012. Los muestreos comenzaron el 2 de julio, dos veces por semana, hasta el momento de la vendimia. Así, se realizaron 15 muestreos para

---

Tempranillo y 18 para Zalema, cuya vendimia es más tardía en la región. Se extrajeron las semillas de cien bayas, tomadas al azar de una muestra bruta de 2 kg. Cada muestra fue a su vez dividida en tres partes de las cuales dos se asignaron al set de calibración del método y la tercera se asignó al set de predicción. Por lo que el número total de muestras fue 99  $((18+15) \times 3)$ .

### 5.3.1. Análisis de referencia

Como análisis de referencia se utilizó una modificación del método espectrofotométrico DMACA (4-(dimetilamino)cinamaldehído) para el análisis de flavanoles desarrollado por Vivas et al. (1994). Cada replicado se separó en dos fracciones, sobre las que se realizaron diferentes extracciones. Una de ellas se llevó a cabo con una disolución metanol:agua (75:25) sobre el polvo obtenido de las semillas liofilizadas y molidas. El resultado de esta extracción se consideró como el total de los flavanoles contenidos en las semillas. En la segunda extracción se realizó una maceración, durante 72 horas, de las semillas enteras y frescas, usando una disolución modelo de acidez, grado alcohólico y fuerza iónica equivalentes a las de un vino estándar (vino sintético), para simular las condiciones en las que las semillas de uva se encuentran durante el proceso de vinificación (Bautista-Ortín et al., 2012).

Los resultados mostraron un descenso en el contenido de flavanoles de las semillas a lo largo de la maduración, independientemente de la variedad de uva y del tipo de extracción realizada (Figura 33). En los extractos con metanol, los valores oscilaron entre 4.28 y 34.26  $\text{mg} \cdot \text{g}^{-1}$  de catequina. Para los extractos con vino sintético, los valores oscilaron entre 0.12 y 7.21  $\text{mg} \cdot \text{g}^{-1}$  de catequina. Se calculó también la extractabilidad, para cada muestra, como el porcentaje de flavanoles que se extraen con vino sintético respecto a los obtenidos mediante la extracción exhaustiva. Estos valores oscilaron ampliamente, aunque se observó un descenso general, siendo aproximadamente un 25% en los primeros estadios y un 5% en el momento de la vendimia.

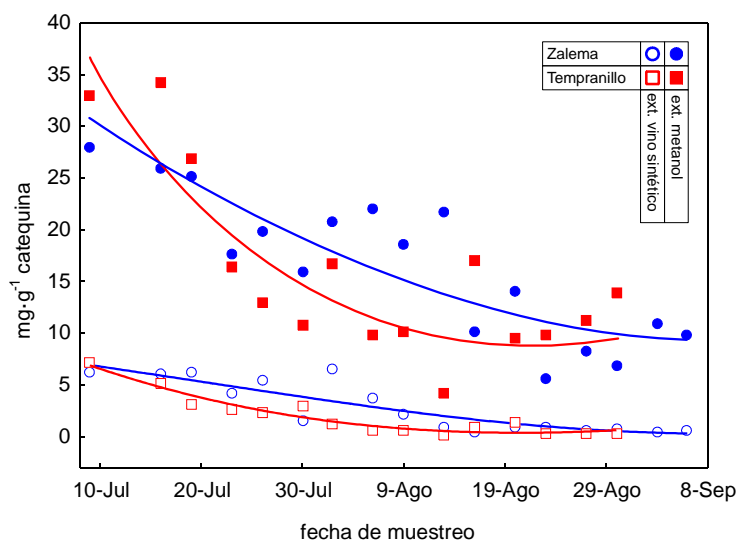


Figura 33. Concentración de flavanoles en los extractos según variedad y tipo de extracción.

### 5.3.2. Adquisición y procesado de imágenes hiperespectrales

El equipo para la adquisición de las imágenes hiperespectrales del Laboratorio Color y Calidad de Alimentos de la Universidad de Sevilla consistió en una cámara Xeva 1.7-320 de Xenics® (Lovaina, Bélgica), un espectrógrafo Inspector N17E de Specim® (Oulu, Finlandia), un dispositivo Mirror Scanner Specim® (Oulu, Finlandia) como sistema de barrido, y dos lámparas halógenas de tipo cuarzo yodo como sistema de iluminación. Las semillas se colocaron sobre una superficie de polietileno, cuyo espectro característico facilitó el proceso de segmentación (Figura 34).



Figura 34. Colocación de semillas en la superficie de polietileno para la adquisición de imágenes.

Para la segmentación se aplicó un método de análisis discriminante por pasos, en lugar de la umbralización utilizada en estudios anteriores, ya que proporcionaba mayor robustez. Para ello, se recopiló una batería representativa de espectros del fondo y de las semillas, que constituyeron las variables independientes en el análisis. Esos espectros se clasificaron según la variable dependiente cualitativa, de pertenencia o no a la región de interés. De las 215 bandas disponibles (950–1650 nm), el análisis seleccionó secuencialmente 3 longitudes de onda para la discriminación de las semillas del fondo: 1216, 1392 y 1147 nm. La Ec. 35 muestra las funciones discriminantes, donde  $R_\lambda$  es la reflectividad del píxel a la longitud de onda  $\lambda$ .

$$\begin{aligned} ROI &= (1430.39 \times R_{1216}) - (1165.10 \times R_{1392}) + (304.26 \times R_{1147}) - 258.90 \\ Fondo &= -(211.98 \times R_{1216}) - (513.85 \times R_{1392}) + (538.86 \times R_{1147}) - 88.56 \end{aligned} \quad \text{Ec. 35}$$

En el algoritmo para la extracción de datos, se aplicaron estas funciones a cada píxel, asignándole la pertenencia o no a la región de interés en función de donde se obtuviera un valor mayor. Para comprobar la idoneidad del método de segmentación, se supervisaron visualmente todas las máscaras de segmentación. A continuación, se consideró el espectro de la muestra como la media de todos los píxeles de la región de interés. A este espectro medio se le aplicó la transformación  $\text{Log}(1/R)$ , el pretratamiento espectral SNV y la segunda derivada. Esta secuencia de

tratamientos fue la que produjo mejores resultados en los análisis estadísticos exploratorios y en las predicciones de la composición flavanólica.

### 5.3.3. Análisis exploratorio y cuantitativo

Preliminarmente, se aplicó un PCA para el estudio de la información latente dentro de la matriz de espectros. El resultado de este análisis no sólo dio información sobre los posibles *outliers* y la distribución de las muestras en el nuevo espacio, sino que también permitió ver la aptitud del set de predicción utilizado en el PLSR posterior. Para la detección de *outliers* se usó el estadístico  $T^2$  de Hotelling, aplicado en las distancias multivariantes de cada muestra respecto al centro del set de datos (Jackson, 2005), el cual, para un nivel de confianza del 95%, eliminó 4 de las 99 muestras según sus características espectrales. A las 95 muestras restantes se les aplicó de nuevo un PCA para asegurar la representatividad del set de predicción dentro del nuevo espacio multivariante generado.

Los PC 1, 2 y 3 explicaron el 98.61%, 1.18% y 0.10% de la varianza, respectivamente. El PC1 estuvo muy afectado por la fecha de muestreo, y sus puntuaciones se incrementaron conforme avanzaba la maduración. Para cada fecha de muestreo, las semillas de Tempranillo siempre tuvieron puntuaciones (*scores*) más altas que las semillas de Zalema. El PC2 mostró diferencias entre las dos variedades de uva. Por lo general, los valores de las semillas de Tempranillo se situaron en el semieje positivo de este PC mientras que los de Zalema lo hicieron en el negativo. Además, se observó que las muestras pertenecientes al set de predicción estuvieron repartidas en el nuevo espacio multivariante de manera homogénea, lo cual demostró su idoneidad para ser usadas como tal.

Para el análisis cuantitativo, se aplicó un PLSR utilizando los resultados del análisis químico como variables dependientes (matriz  $Y$ ) y los datos espectrales procesados como variables independientes (matriz  $X$ ). Para las extracciones con la disolución metanólica, y considerando conjuntamente las muestras de las dos

---

variedades, se obtuvo un coeficiente de determinación  $R^2$  de 0.73 para la calibración y de 0.75 para la predicción. Los errores medios de calibración (RMSEC) y de predicción (RMSEP) fueron 4.01 y 3.86  $\text{mg}\cdot\text{g}^{-1}$ , respectivamente. En el caso de las extracciones con vino sintético, se obtuvieron valores de  $R^2$  de 0.82 para calibración y 0.85 para la predicción, con RMSEC de 0.92  $\text{mg}\cdot\text{g}^{-1}$  y RMSEP de 0.88  $\text{mg}\cdot\text{g}^{-1}$ . Como la variedad resultó ser un factor determinante en el análisis exploratorio preliminar, se decidió repetir el PLSR para cada variedad individualmente. Tal como se esperaba, se observó un aumento de los coeficientes de determinación y un descenso de los errores tanto para la calibración como para la predicción.

Se analizaron también los factores PLS para determinar qué longitudes de onda contribuyeron con mayor peso al modelo. Se observó una alta contribución de la banda entre 1100 y 1300 nm, zona relacionada con las bandas de combinación de las vibraciones de tensión, simétrica y antisimétrica, para el enlace O–H, segundos sobretonos de enlaces C–H aromáticos y terceros sobretonos de C–H (Bokobza, 2007; Ferrer-Gallego et al., 2011). Las bandas de vibración correspondientes a tensiones en el enlace O–H contribuyen al espectro a 1400 nm, por lo que la humedad afecta ampliamente a esta banda. En este caso, esta influencia en los factores PLS puede ser atribuida a la pérdida de agua que sufren las semillas simultánea a la evolución de sus flavanoles durante la maduración (Osborne et al., 1993). Según Goodchild et al. (1998) las bandas cercanas a 1600 nm pueden ser atribuidas a taninos condensados.

#### **5.4. Conclusiones**

- Los modelos de regresión múltiple PLSR permitieron evaluar con éxito el contenido flavanólico de semillas de uva a partir de los espectros de reflectancia en el infrarrojo, con coeficientes de determinación  $R^2$  de 0.75 para cuando la extracción de flavanoides se realizó exhaustivamente a partir del liofilizado de semillas, y de 0.85 para extracciones con semillas

enteras en vino sintético. Además, estos coeficientes alcanzaron valores de hasta 0.88 cuando se consideraron las variedades independientemente.

- Mediante análisis exploratorio PCA fue posible diferenciar las semillas de uva según su variedad partiendo de los datos espectrales. Además, tanto el porcentaje de varianza como la información que ofrecía cada PC coincidieron con los resultados obtenidos en el estudio previo, de puesta a punto de la metodología, donde se usaron diferentes muestras (semillas de otra cosecha) y se realizaron los análisis en un laboratorio diferente. Esto demuestra la robustez de la metodología establecida.
- Las técnicas hiperespectrales, aunque no se postulan aún como sustitutivas, se presentan como una alternativa al análisis químico convencional, debido a su simplicidad y rapidez, ya que es posible evaluar el contenido en flavanoles de las semillas de uva sin necesidad de tratamiento previo de las muestras, y mediante un análisis muy rápido.
- Controlando las variables que afectan a cada variedad de uva, podría llegar a ser un método de referencia para estimar las características químicas de las semillas de uva aparte de una herramienta útil para la industria vitivinícola.



1 **A novel method for evaluating flavanols in grape seeds by near infrared**  
2 **hyperspectral imaging**

3 Rodríguez-Pulido, Francisco J.; Hernández-Hierro, José Miguel; Nogales-  
4 Bueno, Julio; Gordillo, Belén; González-Miret, M. Lourdes.; Heredia, Francisco  
5 J.

6  
7 Food Colour & Quality Lab., Dept. Nutrition & Food Science. Facultad de  
8 Farmacia. Universidad de Sevilla. 41012-Sevilla, Spain.

9

10

11

12 Corresponding author:

13 Heredia, Francisco J.;

14 Food Colour & Quality Lab., Dept. Nutrition & Food Science. Facultad de  
15 Farmacia. Universidad de Sevilla. 41012-Sevilla, Spain

16 Tel.: +34 954556495 Fax: +34 954556110

17 e-mail: heredia@us.es

18

19 **ABSTRACT**

20 Chemical composition of seeds changes during grape ripening and this affects  
21 the sensory properties of wine. In order to control the features of wines, the  
22 condition of seeds is becoming an important factor for deciding the moment of  
23 harvesting by winemakers. Sensory analysis is not easy to carry out and  
24 chemical analysis needs lengthy procedures, reagents, and it is destructive and  
25 time-consuming. In the present work, near infrared hiperespectral imaging has  
26 been used to determine flavanols in seeds of red (cv. Tempranillo) and white  
27 (cv. Zalema) grapes (*Vitis vinifera* L.). As reference measurements, the flavanol  
28 content was estimated using the *p*-dimethylaminocinnamaldehyde (DMACA)  
29 method. Not only total flavanol content was evaluated but also the quantity of  
30 flavanols that would be extracted into the wine during winemaking. A like-wine  
31 model solution was used for this purpose. Calibrations were performed by  
32 partial least squares regression and they provide coefficients of determination  
33  $R^2=0.73$  for total flavanol content and  $R^2=0.85$  for predicting flavanols extracted  
34 with model solution. Values up to  $R^2=0.88$  were reached when cultivars were  
35 considered individually.

36

37 **KEYWORDS**

38 Chemometrics; Flavanols; Grape seeds; Hyperspectral imaging; Near infrared;  
39 *Vitis vinifera*;

40

## 41 **1. Introduction**

42 Grape seeds constitute a small part of the berry, but they affect extensively the  
43 sensory properties of wine. Their phenolic compounds are responsible of these  
44 properties and they change in a qualitative and quantitative manner during  
45 ripening [1,2]. The most representative of them in grape seeds, flavanols,  
46 include flavan-3-ol monomers (catechin, epicatechin and epicatechin gallate)  
47 and procyanidins, which are polymers comprised of flavan-3-ol terminal and  
48 extension subunits [3]. Another aspect that has raised interest is the  
49 extractability of these compounds. It has been reported that extractability  
50 depends on the ripeness of grape seeds. This phenomenon is due to changes  
51 in the interactions between tannins and cell wall material [4]. Insufficiently  
52 ripened grapes have a lower tannin extractability [5]. Since changes during  
53 ripening affect both gustatory and appearance properties, sensorial analysis is  
54 the most common approach to evaluate the condition of the seeds by vine  
55 growers, though it is difficult to be carried out in an accurate and objective  
56 manner [6]. Chemical analysis is the most widely accepted reference method  
57 for analysing seeds. Nevertheless, these methods frequently are destructive,  
58 time-consuming, and entail the use of reagents [7-10]. Replacing conventional  
59 analyses, near infrared (NIR) spectroscopy provides fast, accurate and non-  
60 destructive way to obtain chemical composition [11,12]. These techniques have  
61 been successfully joined to computer vision systems [13,14]. These systems  
62 have the limitation of mainly acting on the surface of materials, thus principally  
63 allowing identification of external features. However, phenolic compounds are  
64 mainly concentrated within the outer layer of grape seeds [15].

65 Near-infrared (NIR) hyperspectral imaging is a powerful technique which has  
66 been used in several applications in agricultural products [16-20]. In fact, it has  
67 been applied to grape seeds for establishing the methodology for acquiring  
68 images, discriminating varieties and estimating the date of sampling, but not yet  
69 for predicting chemical composition [21]. Hyperspectral imaging provides a  
70 digital image and the spectrum belonging to each pixel. Hyperspectral images  
71 (HSI), or hypercubes, are three-dimensional data matrix where the first two  
72 axes of the matrix represent the spatial coordinates, while the third axis portrays  
73 the spectral dimension. They usually are represented as a battery of images  
74 where each layer shows the reflectance at a wavelength in grey scale [22]. Due  
75 to the great amount of information that they include, HSI require the application  
76 of multivariate data analysis for data exploration. As with NIR spectroscopy,  
77 chemometric techniques are applied to decompose the image dataset, process  
78 and perform regression or classification analyses. The possibilities of  
79 hyperspectral imaging based on the NIR range have been illustrated developing  
80 a model able to predict and classify barley kernels [23,24], predicting hardness  
81 in maize kernels [25], and studying enzymes activity and detecting sprout  
82 damage in wheat [26,27].

83 The most frequently used methods for spectral correction are multiplicative  
84 scatter correction (MSC), standard normal variate (SNV) and derivation. [28-31].  
85 However, there is still no standard procedure to decide which spectral  
86 processing produce best results. Partial least squares regression (PLSR) is a  
87 procedure used to relate a large number of independent variables (predictors)  
88 to one (PLSR1) or few (PLRS2) response variables (observations) when a

89 reduced number of cases are available. Since it reduces a great number of  
90 redundant information, it is very effective in spectral analysis [32,33].

91 The aim of this work was to evaluate the potential of NIR hyperspectral imaging  
92 for the evaluation of flavanols in seeds from red and white grapes during  
93 ripening.

94

## 95 **2. Material and methods**

### 96 **2.1 Sampling**

97 The grapes (*Vitis vinifera* L.) sampled are included under the “Condado de  
98 Huelva” Designation of Origin, in Southwestern Spain, harvested in 2012. One  
99 red variety (cv. Tempranillo) and one autochthonous white variety (cv. Zalema)  
100 were used. The number of samples was 18 for Zalema and 15 for Tempranillo,  
101 depending on the availability and harvesting times of each variety. They were  
102 taken twice a week from early July until postharvest mid-September. Sampling  
103 process was carried out at daybreak by taking a pair of berries from alternate  
104 grapevines, from four rows of vines, and from both sides of each row up to  
105 reach 2 kg of berries. In this process, the berries were taken with pedicel intact  
106 to slow down the berry oxidation as long as possible. Once in laboratory, one  
107 hundred berries were randomly taken and seeds removed, left to dry at room  
108 temperature for 2 hours, and frozen at  $-20^{\circ}\text{C}$  until acquisition of hyperspectral  
109 images and chemical analysis. Each sample was divided into three parts used  
110 as replicates  $((18+15)\times 3=99$  samples). Two of these replicates were allocated  
111 to the calibration set and the other sample to the prediction set.

112

113 2.2 Hyperspectral image analysis

114 The system comprised a Xenics<sup>®</sup> XEVA-USB InGaAs camera (320 × 256 pixels;  
115 Xenics Infrared Solutions, Inc., Leuven, Belgium), a spectrograph (Specim  
116 ImSpector N17E Enhanced; Spectral Imaging Ltd., Oulu, Finland) covering the  
117 spectral range between 884 and 1717 nm (spectral resolution of 3.25 nm), two  
118 70W tungsten iodine halogen lamps (Prilux<sup>®</sup>, Barcelona, Spain) used as light  
119 source, a mirror scanner (Spectral Imaging Ltd., Oulu, Finland), and a computer  
120 system. HSI were recorded using a 50 Hz frame rate and an exposure time of 9  
121 ms using the instrument acquisition software SpectralDAQ 3.62 (Spectral  
122 Imaging Ltd., Oulu, Finland). From the acquired HSI, it was observed that the  
123 first and the last twenty bands of the image had a high level of noise, thus not  
124 being useful for spectral data extraction. Therefore, images were cropped to the  
125 spectral range of 950-1650 nm with a total of 215 bands.

126 A 'white reference' image ( $W$ , 100% reflectance) was acquired from a white  
127 Spectralon<sup>®</sup> ceramic tile (Labsphere Inc., North Sutton, USA), and a 'dark  
128 reference' image ( $B$ , 0% reflectance) was obtained with the light source off and  
129 the mirror scanner completely covered with its opaque cap. The white and dark  
130 'reference' HSI were used to correct the raw images ( $R_0$ ) in order to obtain a  
131 relative reflectance image ( $R$ ) according to the following equation:

132 
$$R = \frac{R_0 - B}{W - B} \quad (1)$$

133 For segmentation of HSI, a novel method based on forward stepwise  
134 discriminant analysis was applied with the software Statistica 8.0 [34]. Image  
135 processing, spectral processing and statistical treatment were carried out using

136 MATLAB R2012b [35]. A flowchart of the image processing and spectral  
137 treatment employed in this study is schematized in Figure 1.

138

### 139 2.3 Chemical analysis

140 Each sample was split into two fractions subjected to different extractions. For  
141 the exhaustive extraction, grape seeds were freeze-dried and ground to obtain  
142 a homogeneous powder for extraction. One gram of seed powder was extracted  
143 with ten millilitres of methanol:water (75:25), sonicated (15 minutes) and  
144 centrifuged (15 minutes), repeating the extraction process twice more. The  
145 methanolic extracts were combined and finally made up to 50 mL with  
146 methanol. For the extraction in wine-like medium, two grams of intact grape  
147 seeds were macerated in 50 mL of model wine solution (4 g·L<sup>-1</sup> tartaric acid,  
148 12.5% ethanol, adjusted at pH 3.6 with NaOH 0.5 M) during 72 h [36].

149 Flavanols spectrophotometric analysis of both extractions was carried out  
150 following a modification of Vivas et al. [37]. Ten or twenty microlitres of total  
151 extraction or wine like medium extracts were mixed with 190 or 180 µL of  
152 methanol respectively and 1 mL of DMACA reagent. The DMACA reagent was  
153 prepared immediately before use, containing 0.1% (w/v) DMACA (4-  
154 dimethylaminocinnamaldehyde) in a mixture of HCl:methanol (1:10, v/v). The  
155 analyses were performed in triplicate on an Agilent 8453 UV-visible  
156 spectrophotometer (Palo Alto, USA), equipped with diode array detection  
157 (DAD), measuring absorbance at 640 nm and using a calibration curve of (+)-  
158 catechin (Sigma-Aldrich, St. Louis, USA) for quantification. The aforesaid  
159 extract volumes were appropriately modified when the concentration was outside

160 the lineal range of the calibration curve. All results were expressed as mg of  
161 catechin equivalents per gram of grape seed.

162

### 163 **3. Results and discussion**

#### 164 3.1 Segmentation by discriminant analysis

165 A set of reflectance spectra belonging to seeds and background was collected  
166 as input data set. The forward stepwise discriminant analysis included  
167 sequentially three wavelengths, 1216, 1392, and 1147 nm for discriminating the  
168 region of interest from the background. Figure 2 shows the average spectra  
169 belonging to seeds and background (a homogeneous surface composed of  
170 polyethylene) and highlights the selected bands. The algorithm of segmentation  
171 saved all the masks of segmentation and they were visually supervised for  
172 ensuring the suitability of the proposed method.

173

#### 174 3.2 Exploratory Analysis of Spectra

175 Figure 3 shows the mean and standard deviation spectra regarding the variety  
176 of grape seeds. It also shows the spectra after applying the transformations  
177  $\text{Log}(1/R)$ , SNV treatment, and second derivative, treatments that yielded the  
178 best results in prediction analyses. It can be seen that seeds from white and red  
179 grapes have different reflectance intensities along some wavelength regions,  
180 although with the same pattern.

181 Before the quantitative analysis, principal component analysis (PCA) was used  
182 as unsupervised pattern recognition technique in order to get information about  
183 the latent structure of the spectral matrix. This method provided not only



184 information related to spectral outliers and the distribution of samples in the  
185 newly-created space but also was an important source of knowledge with which  
186 to evaluate the suitability of prediction set used in PLSR. For detecting possible  
187 outliers, Hotelling's  $T^2$  statistic was used as a measure of the multivariate  
188 distance of each sample from the centre of the data set [38]. Regarding the  
189 spectral features of each sample, this test rejected 4 of the 99 samples  
190 considering a confidence level of 95%. Using the spectral data of the remaining  
191 samples (without outliers), PCA was applied again in order to ensure the  
192 representativeness of the prediction set in the generated multivariate space.  
193 PC1, PC2, and PC3 explained 98.61%, 1.18%, and 0.10% of the total variance  
194 respectively. PC1 was influenced by the time in an extensively manner. Figure  
195 4a shows PC1 and its evolution over time. At every date, spectra from  
196 Tempranillo seeds had higher scores than Zalema ones. Moreover, this  
197 dependency seemed stronger for Tempranillo, being its slope higher. Figure 4b  
198 shows the scatterplot of scores for PC2 and PC3. Generally, Tempranillo seeds  
199 presented positive scores for PC2 while Zalema seeds presented negative  
200 ones. Furthermore, it can be observed that samples belonging to prediction set  
201 were uniformly distributed among calibration set samples. These results are in  
202 agreement with results previously reported [21].

203

### 204 3.3 Quantitative analysis

205 Flavanols content decreased during the grape ripening regardless the variety  
206 and type of extraction. The methanol extract flavanols ranged from 4.28 to  
207 34.26  $\text{mg}\cdot\text{g}^{-1}$  of grape seed. The flavanols from the extracts obtained using like-

208 wine solution ranged from 0.12 to 7.21 mg·g<sup>-1</sup> of grape seed. Table 1 shows a  
209 brief resume of the aforementioned results. It must be highlighted that high  
210 standard deviations were due to the evolution during ripeness instead of errors  
211 of measurements. Although it was not the goal of this work, extractability of  
212 each sample was evaluated as the fraction of flavanols extracted by the model  
213 solution with respect to the exhaustive extraction. The extractability also  
214 decreased during ripening, being about 25% at the first stages and about 5% in  
215 the last ones.

216 Results of chemical analyses were used as dependent (Y) variables and the  
217 matrix of processed spectra was used as the independent (X) variables in the  
218 PLSR. The statistical parameters of the final calibration equations are shown in  
219 Table 2. For extractions with methanol and considering all samples as a unique  
220 data set, R<sup>2</sup> was 0.73 for calibration and 0.75 for prediction. The RMSEC and  
221 RMSEP were 4.01 and 3.86 mg·g<sup>-1</sup> of grape seed respectively. Results for  
222 predicting flavanols extracted by like-wine solution had R<sup>2</sup>=0.82 for calibration  
223 and R<sup>2</sup>=0.85 for prediction. In this case, RMSEC and RMSEP were 0.92 and  
224 0.88 mg·g<sup>-1</sup> of grape seed respectively. Since cultivar was a determining factor  
225 in preliminary exploratory analysis, the PLSR were repeated for each variety  
226 individually. Because of this, results in Tables 1 and 2 are also broken down  
227 into varieties. As it was expected, coefficients of determination increased while  
228 RMSEC and RMSEP decreased.

229 Figure 5 shows the loadings resulting of the PLSR model for total flavanols and  
230 it indicates the most dominant wavelengths. The spectral region between 1100-  
231 1300 nm showed important contribution to the model loadings and is mainly

232 related to the combination band of O-H symmetric and anti-symmetric stretching  
233 vibration, the combination band of C-H aromatic second overtones, and C-H  
234 third overtone vibration. These can be attributed to the chemical structure of  
235 phenolic compounds [39,40]. The first O-H stretching overtone contributes to  
236 spectrum at 1400 nm, hence the moisture affects expansively to this band. In  
237 this case, the influence can be attributed to the loss of water that grape seeds  
238 suffer at the same time that flavanols develop [41]. According to Goodchild *et al.*  
239 [42], bands close to 1600 nm are attributed to condensed tannins.

240

#### 241 **4. Conclusions**

242 The PLSR models were successfully performed to evaluate flavanols in grape  
243 seeds. These were able to predict the concentration of flavanols of a sample  
244 based on spectral features as the predictor variables with a coefficient of  
245 determination of  $R^2$  of 0.75 for total extractions and 0.85 for extractions with  
246 model wine solution. Furthermore, this coefficient reached up to 0.88 when  
247 varieties were considered individually. On the other hand, PCA was suitable for  
248 grape seeds characterization regarding the variety, proving the suitability of the  
249 methodology previously established.

250 This work has shown the capability of hyperspectral imaging to evaluate the  
251 flavanol content in grape seeds without any sample treatment. Though it is not  
252 yet a substitute for conventional chemical analysis, it is an attractive alternative  
253 due to its simplicity and quickness. By establishing the variables that affects  
254 each cultivar, this could become a reference method to assess the chemical

255 characteristics of grape seeds during maturation, being very useful for vine  
256 growers and wineries.

257

### 258 **Acknowledgements**

259 This work was supported by the projects P10-AGR6331 (Consejería de  
260 Economía, Innovación, Ciencia y Empresa, Junta de Andalucía), AGL2011-  
261 30254-C02 (Ministerio de Economía y Competitividad, Gobierno de España).

262 The Spanish MICINN is also thanked for F.J. Rodríguez-Pulido, J. Nogales-  
263 Bueno FPI grants (BES-2009-025429 and BES-2012-060192 respectively) and  
264 J.M. Hernández-Hierro 'Juan de la Cierva' contract (JCI-2011-09201).

265

266 **References**

- 267 [1] Y. Cadot, M.T. Miñana-Castelló, and M. Chevalier, Journal of  
268 Agricultural and Food Chemistry, 54 (2006) 9206-9215.
- 269 [2] B.S. Sun, T. Pinto, M.C. Leandro, J.M. Ricardo-Da-Silva, and M.I.  
270 Spranger, Am. J. Enol. Vitic., 50 (1999) 179-184.
- 271 [3] R. Ristic and P.G. Iland, Australian Journal of Grape and Wine  
272 Research, 11 (2005) 43-58.
- 273 [4] R.L. Hanlin, M. Hrmova, J.F. Harbertson, And M.O. Downey, Australian  
274 Journal of Grape and Wine Research, 16 (2010) 173-188.
- 275 [5] C. Peyrot des Gachons and J.A. Kennedy, Journal of Agricultural and  
276 Food Chemistry, 51 (2003) 5877-5881.
- 277 [6] J. Rousseau and D. Delteil, Revue française d'oenologie, 183 (2000) 10-  
278 13.
- 279 [7] A.K. Sandhu and L. Gu, Journal of Agricultural and Food Chemistry, 58  
280 (2010) 4681-4692.
- 281 [8] T. Fuleki and J.M. Ricardo da Silva, Journal of Agricultural and Food  
282 Chemistry, 45 (1997) 1156-1160.
- 283 [9] J.A. Kennedy, M.A. Matthews, and A.L. Waterhouse, Phytochemistry, 55  
284 (2000) 77-85.
- 285 [10] E. Obreque-Slier, A. Peña-Neira, R. López-Solís, F. Zamora-Marín, J.M.  
286 Ricardo-da Silva, and O. Laureano, Journal of Agricultural and Food  
287 Chemistry, 58 (2010) 3591-3599.
- 288 [11] B.M. Nicolaï, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron,  
289 and J. Lammertyn, Postharvest Biology and Technology, 46 (2007) 99-118.

- 290 [12] J.M. Hernández-Hierro, J. Valverde, S. Villacreces, K. Reilly, M. Gaffney,  
291 M.L. González-Miret, F.J. Heredia, and G. Downey, *Journal of Agricultural and*  
292 *Food Chemistry*, 60 (2012) 7352-7358.
- 293 [13] D.F. Barbin, G. ElMasry, D.W. Sun, and P. Allen, *Food Chemistry*, 138  
294 (2013) 1162-1171.
- 295 [14] D.W. Sun, *Journal of Food Engineering*, 61 (2004) 1-2.
- 296 [15] J.H. Thorngate and V.L. Singleton, *Am. J. Enol. Vitic.*, 45 (1994) 259-  
297 262.
- 298 [16] T.M. Baye, T.C. Pearson, and A.M. Settles, *Journal of Cereal Science*,  
299 43 (2006) 236-243.
- 300 [17] M.A. Shahin and S.J. Symons, *Computers and Electronics in*  
301 *Agriculture*, 75 (2011) 107-112.
- 302 [18] S. Cubero, N. Aleixos, E. Moltó, J. Gómez-Sanchis, and J. Blasco, *Food*  
303 *Bioprocess Technol*, 4 (2011) 487-504.
- 304 [19] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O.L. García-  
305 Navarrete, and J. Blasco, *Food Bioprocess Technol*, 5 (2012) 1121-1142.
- 306 [20] A. Baiano, C. Terracone, G. Peri, and R. Romaniello, *Computers and*  
307 *Electronics in Agriculture*, 87 (2012) 142-151.
- 308 [21] F.J. Rodríguez-Pulido, D.F. Barbin, D.W. Sun, B. Gordillo, M.L.  
309 González-Miret, and F.J. Heredia, *Postharvest Biology and Technology*, 76  
310 (2013) 74-82.
- 311 [22] J. Burger and P. Geladi, *Analyst*, 131 (2006) 1152-1160.
- 312 [23] L. Munck and B. Møller, *Journal of the Institute of Brewing*, 110 (2004)  
313 3-17.

- 314 [24] P. Engelbrecht, M. Manley, P.J. Williams, G.D. Toit, and P. Geladi, Pre-  
315 germination detected in whole cereal grains using near infrared hyperspectral  
316 imaging, *Proceedings of the CST SA - ICC International Grains Symposium*,  
317 Quality and Safety of Grain Crops and Foods, 2010, pp. 123-127.
- 318 [25] P. Williams, P. Geladi, G. Fox, and M. Manley, *Analytica Chimica Acta*,  
319 653 (2009) 121-130.
- 320 [26] J. Xing, S. Symons, M. Shahin, and D. Hatcher, *Biosystems*  
321 Engineering, 106 (2010) 188-194.
- 322 [27] J. Xing, P. Van Hung, S. Symons, M. Shahin, and D. Hatcher, *Sensing*  
323 and Instrumentation for Food Quality and Safety, 3 (2009) 211-218.
- 324 [28] P. Geladi, D. MacDougall, and H. Martens, *Appl. Spectrosc.*, 39 (1985)  
325 491-500.
- 326 [29] T. Isaksson and T. Næs, *Appl. Spectrosc.*, 42 (1988) 1273-1284.
- 327 [30] M. Kaihara, T. Takahashi, T. Akazawa, T. Sato, and S. Takahashi,  
328 *Spectroscopy Letters*, 35 (2002) 369-376.
- 329 [31] C. Pizarro, I. Esteban-Díez, A.J. Nistal, and J.-M. González-Sáiz,  
330 *Analytica Chimica Acta*, 509 (2004) 217-227.
- 331 [32] P. Lin, Y. Chen, and Y. He, *Food and Bioprocess Technology*, 5 (2012)  
332 235-242.
- 333 [33] M.M. Pojic and J.S. Mastilovic, *Food and Bioprocess Technology*, 6  
334 (2013) 330-352.
- 335 [34] StatSoft Inc.. *Statistica 8.0*. 2007. Tulsa, USA, StatSoft Inc.
- 336 [35] The Mathworks. *MATLAB R2012b*. 2012. Natick, USA, The MathWorks  
337 Inc.

338 [36] A.B. Bautista-Ortín, P. Rodríguez-Rodríguez, R. Gil-Muñoz, E. Juménez-  
339 Pascual, N. Busse-Valverde, A. Martínez-Cutillas, J.M. López-Roca, and E.  
340 Gómez-Plaza, *Australian Journal of Grape and Wine Research*, 18 (2012) 123-  
341 130.

342 [37] N. Vivas, Y. Glories, L. Lagune, C. Saucier, and M. Augustin, *Journal*  
343 *International des Sciences de la Vigne et du Vin*, 28 (1994) 319-336.

344 [38] J.E. Jackson, *A User's Guide to Principal Components*, Wiley, 2005.

345 [39] L. Bokobza, in: H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise  
346 (Eds.), *Near-Infrared Spectroscopy: Principles, Instruments, Applications*,  
347 Wiley-VCH Verlag GmbH, Weinheim, Germany, 2007, pp. 11-41.

348 [40] R. Ferrer-Gallego, J.M. Hernández-Hierro, J.C. Rivas-Gonzalo, and M.T.  
349 Escribano-Bailón, *LWT - Food Science and Technology*, 44 (2011) 847-853.

350 [41] B.G. Osborne, T. Fearn, and P.T. Hindle, *Practical NIR Spectroscopy*  
351 *With Applications in Food and Beverage Analysis*, Longman Group, United  
352 Kingdom 1993.

353 [42] A.V. Goodchild, F.J. El Haremein, A. Abd El Moneim, H.P.S. Makkar, and  
354 P.C. Williams, *Journal of Near Infrared Spectroscopy*, 6 (1998) 175-181.

355



356 **FIGURE CAPTIONS**

357 Figure 1. Flow chart of the image processing and spectral treatment employed  
358 in this study.

359 Figure 2. Spectra of seeds and background highlighting the bands included by  
360 the forward stepwise discriminant analysis.

361 Figure 3. Average reflectance spectra and average processed spectra of each  
362 variety. Shaded areas represent the standard deviation at each wavelength.

363 Figure 4. (a) Dependency of PC1 with date. (b) Scatterplot of scores for PC2  
364 and PC3. Circles represent Zalema and squares represent Tempranillo, in turn,  
365 filled and unfilled marks belong to calibration and prediction sets respectively.

366 Figure 5. Loadings plot for the first three PLS Factors of the regression model  
367 for total flavanols prediction.

**Table 1**

Table 1. Summary of chemical analyses for all samples and regarding the variety (all results were expressed as mg of catechin equivalents per gram of grape seed).

	N	Extraction	Mean	Minimum	Maximum	Std.Dev.
All Samples	95	model wine	2.26	0.12	7.21	2.22
		total	15.82	4.28	34.26	7.74
Zalema	50	model wine	2.54	0.38	6.48	2.40
		total	15.85	5.63	28.05	6.93
Tempranillo	45	model wine	1.95	0.12	7.21	1.98
		total	15.78	4.28	34.26	8.63

**Table 2**

Table 2. Calibration and prediction results for the PLS models obtained from processed spectra (all results were expressed as mg of catechin equivalents per gram of grape seed).

	N	Extraction	PLS Factors	$R^2_c$	RMSEC	$R^2_p$	RMSEP
All samples	95	model wine	3	0.82	0.92	0.85	0.88
		total	3	0.73	4.01	0.75	3.86
Zalema	50	model wine	2	0.83	0.98	0.85	0.92
		total	1	0.82	2.90	0.82	2.93
Tempranillo	45	model wine	2	0.88	0.67	0.88	0.69
		total	6	0.94	2.09	0.88	2.89

Figure 1

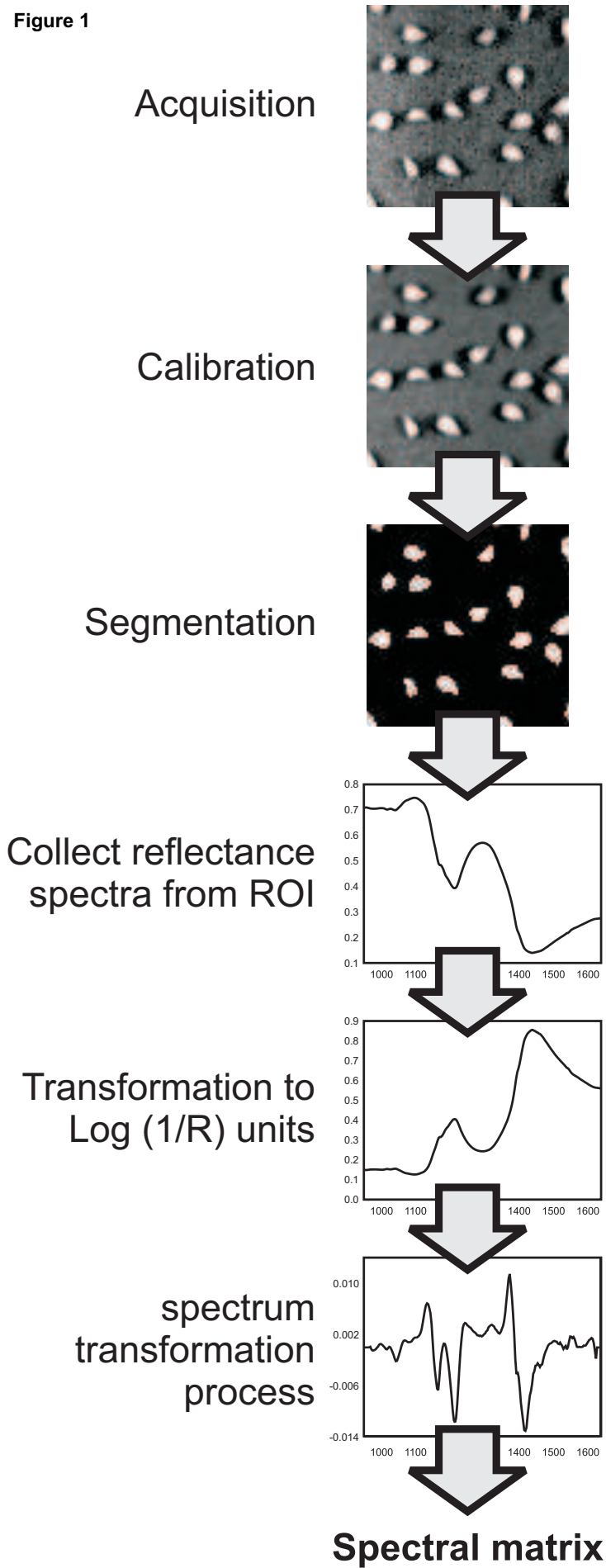


Figure 2

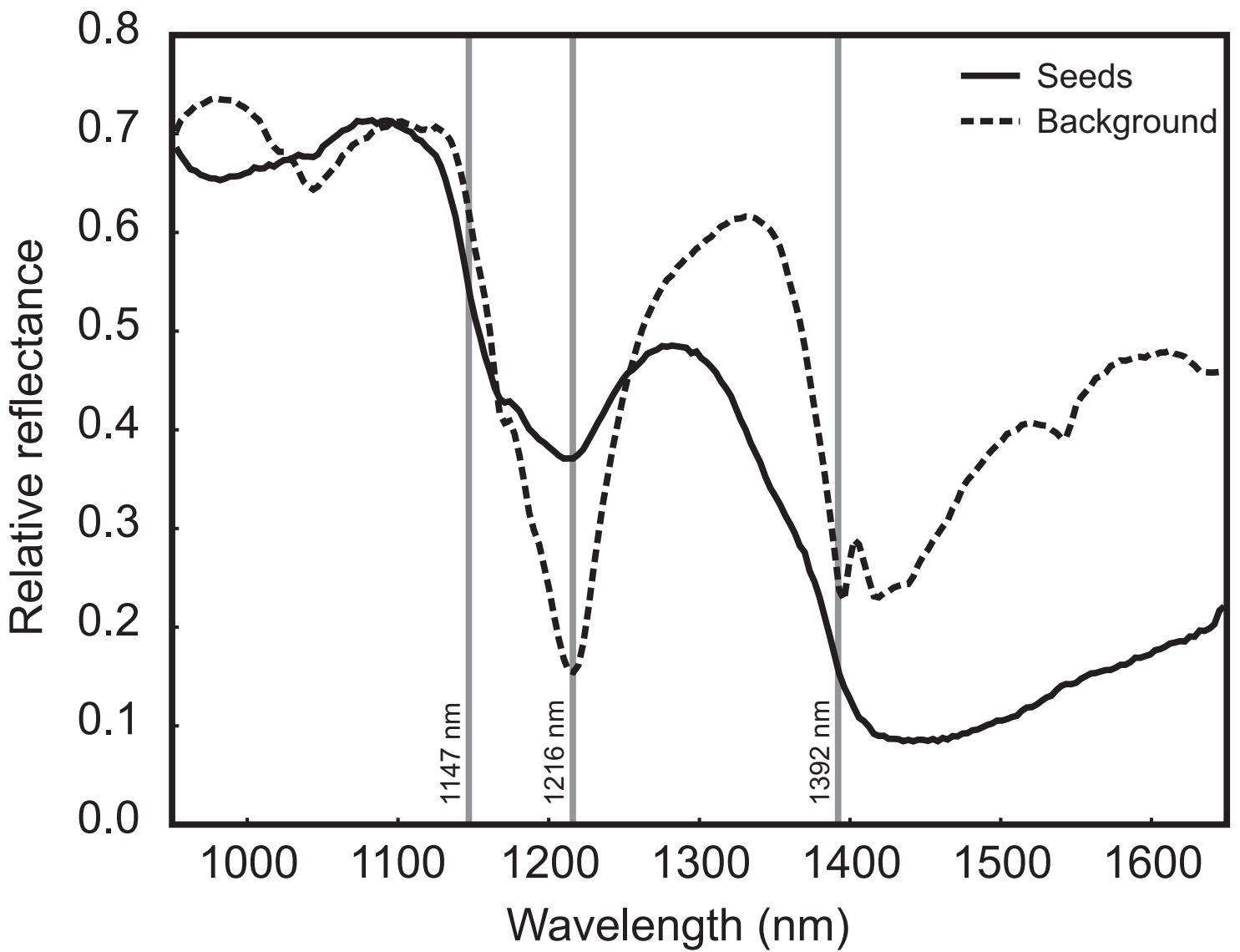


Figure 3

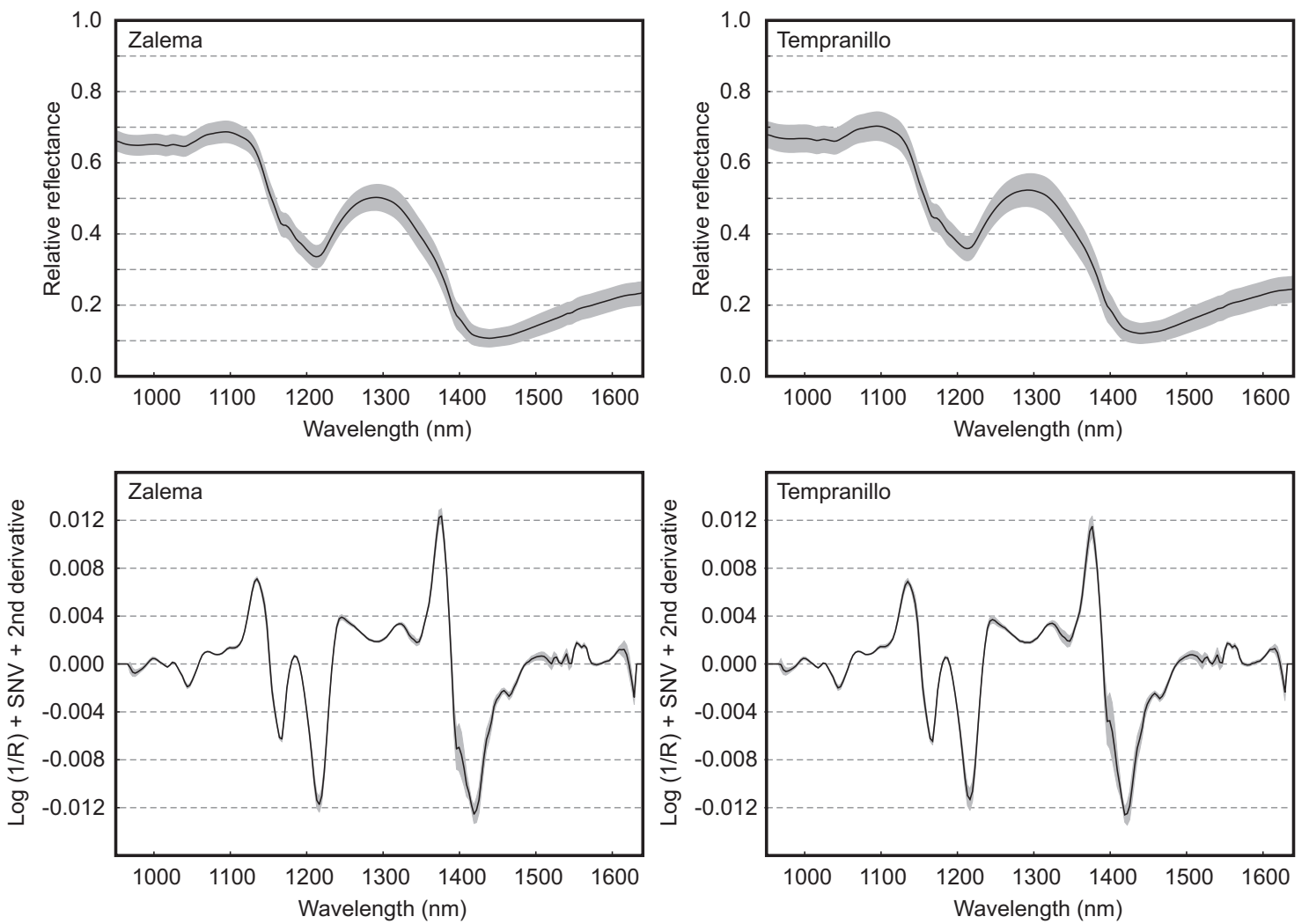


Figure 4

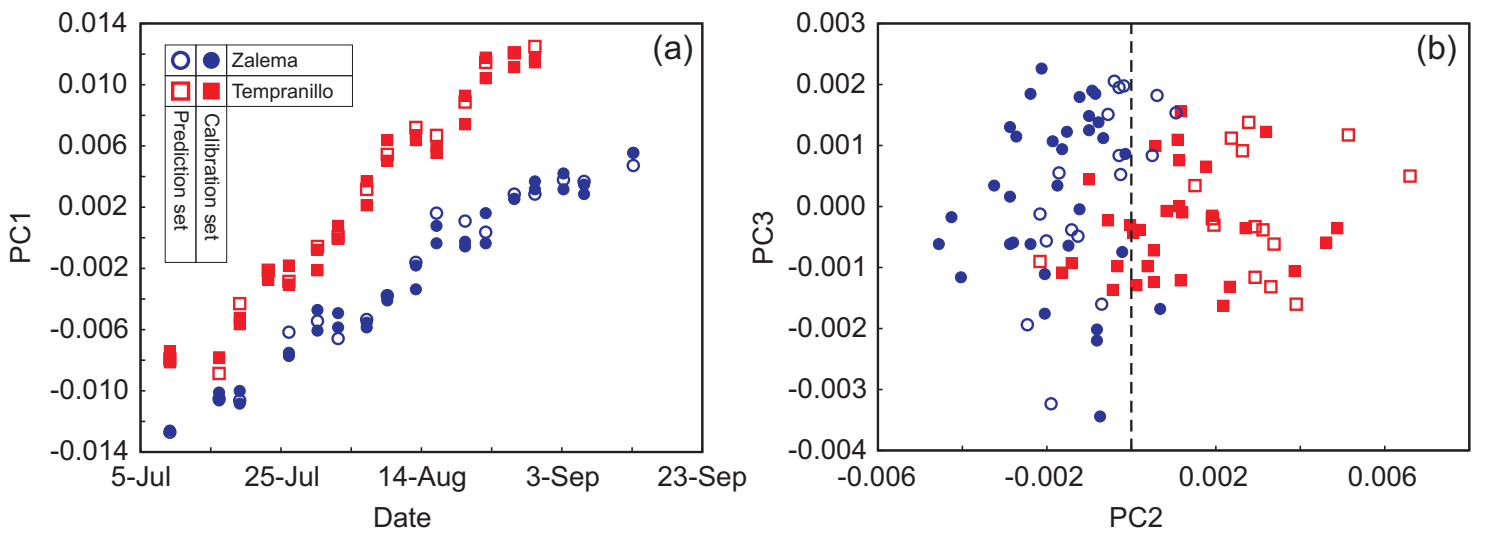
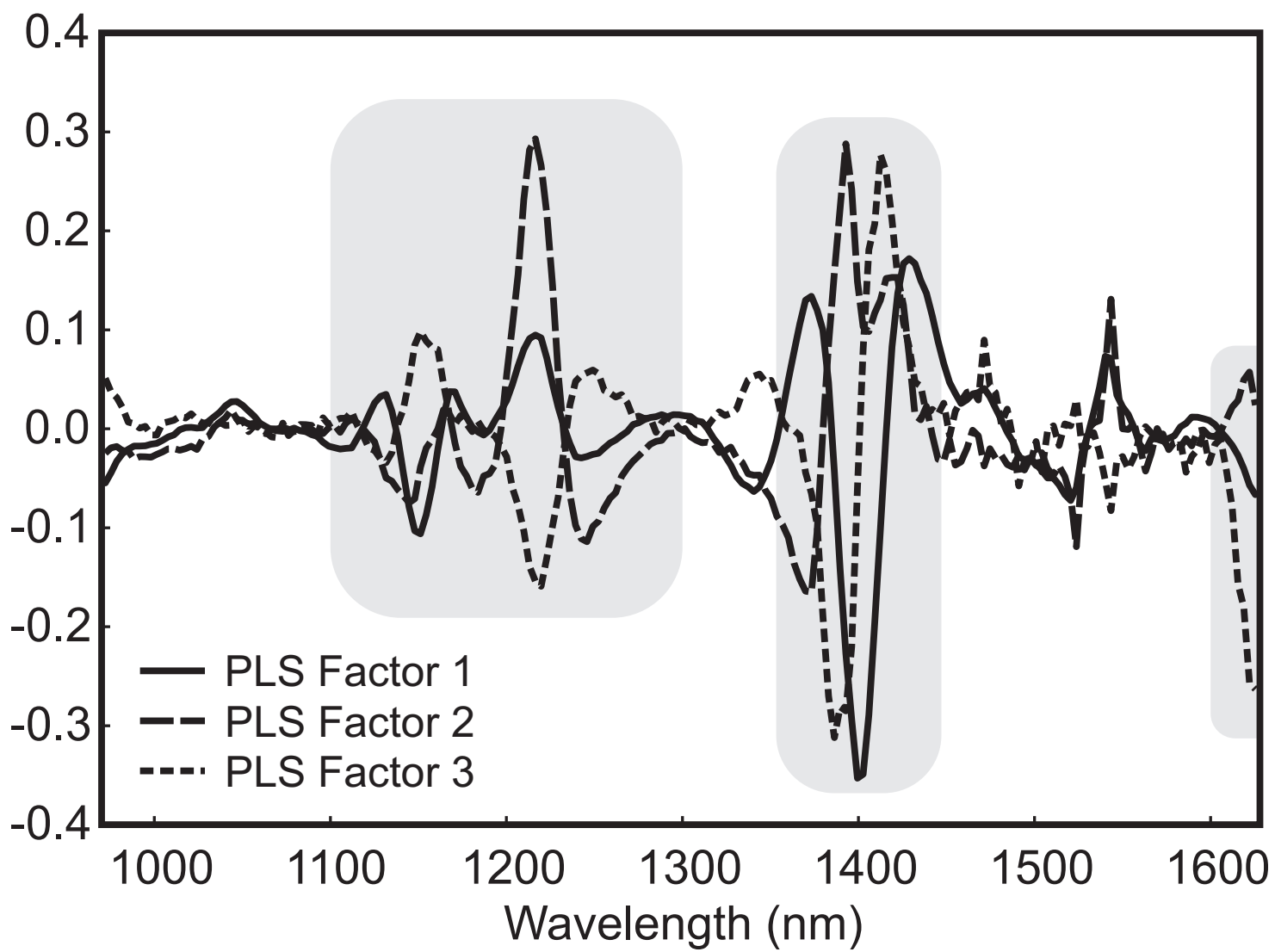


Figure 5







## ***CONCLUSIONES GENERALES***



---

## CONCLUSIONES GENERALES

Las conclusiones más importantes que se deducen de los resultados obtenidos son las siguientes:

PRIMERA. El análisis de imagen resulta una técnica útil para evaluar objetivamente las características morfológicas y colorimétricas de las semillas de uva. Mediante la aplicación de esta técnica se ha observado un notable descenso de las coordenadas claridad y croma durante la maduración. Por su parte, las características morfológicas obtenidas a partir de la imagen han resultado ser útiles para la discriminación varietal, siendo la relación de aspecto y el grado de redondez los parámetros más útiles en la dicha discriminación.

SEGUNDA. Se ha propuesto un índice de pardeamiento basado en coordenadas colorimétricas para el seguimiento y determinación del estado de maduración de las semillas de uva. Este método ha resultado óptimo para evaluar el avance de la maduración, y su idoneidad ha sido probada en muestras de semilla de diferentes variedades de uvas blancas y tintas, en dos cosechas diferentes.

TERCERA. Se ha desarrollado un método para la evaluación de la heterogeneidad del color, basado en la división de las regiones de color según el umbral de percepción del ojo humano, y en el ajuste de la nube de puntos de cada grupo a un elipsoide en el espacio CIELAB, cuya orientación, tamaño y posición depende de las características colorimétricas de la muestra. Este método supone un avance en el modo de expresar y cuantificar la heterogeneidad cromática en alimentos.

CUARTA. La posición, dimensiones y orientación de los elipsoides, creados a partir de las nubes de puntos de color, han resultado eficaces para la predicción de propiedades químicas de uvas blancas. Estas características también se han aplicado con éxito a la clasificación de semillas en diferentes estadios de maduración dentro de una muestra. El método ha permitido encontrar

---

patrones de evolución característicos en la maduración de las semillas para las semillas de diferentes variedades de uva.

QUINTA. Las características de apariencia obtenidas por técnicas de imagen resultaron ser útiles para la evaluación de la composición química de las semillas. Así, se ha comprobado la alta correlación entre la madurez fenólica de las semillas de uva y sus coordenadas colorimétricas mediante la aplicación de técnicas estadísticas multivariantes.

SEXTA. Se ha comprobado la eficacia del análisis de imágenes hiperespectrales como técnica para la evaluación de las semillas de uva. A partir de los espectros de reflectancia, se ha podido discriminar entre variedades de semillas y diferenciar factores agronómicos como el tipo de suelo.

SÉPTIMA. Se ha probado la eficacia de la reducción de variables mediante el análisis de componentes principales para la discriminación varietal. Los resultados obtenidos a partir de las bandas seleccionadas fueron comparables a los obtenidos a partir del espectro infrarrojo cercano completo. Además, las bandas seleccionadas coincidieron con las longitudes de onda que, en estudios anteriores, resultaron ser las que aportaban la información química más relevante.

OCTAVA. Mediante la aplicación de las herramientas quimiométricas adecuadas, y a partir de los espectros de reflectancia en la región del infrarrojo, es posible estimar el contenido flavanólico de las semillas de uva. Esta capacidad predictiva mejora cuando, en lugar de evaluar el contenido total, se considera solo la cantidad de flavanoles que se transfieren al mosto durante el proceso de vinificación.

NOVENA. Las técnicas de análisis de imagen desarrolladas, tanto en la región visible del espectro como en la de infrarrojo cercano se presentan como una alternativa para la estimación de características químicas de las semillas, debido a su simplicidad, versatilidad, rapidez y bajo coste. No necesitan tratamiento previo de las muestras ni uso de reactivos, lo cual les hace

---

potencialmente útiles para la estimación rápida del estado de madurez de las semillas de uva en la industria vitivinícola.

---

---

## GENERAL CONCLUSIONS

The following main conclusions are drawn from the studies carried out.

FIRST. Image analysis is a useful tool for evaluating the morphological and colorimetric characteristics of grape seeds. A noticeable decrease of the coordinates lightness and chroma was observed during the ripening by means of this technique. In turn, the morphological features obtained were useful for varietal discrimination, being aspect ratio and roundness the most useful parameters in that discrimination.

SECOND. A browning index based on the colorimetric coordinates has been proposed for monitoring and determining the ripening stage of grape seeds. This method was optimal for assessing the ripening evolution, and its suitability was proved in samples of seeds from white and red grape varieties, belonging to two vintages.

THIRD. A method for evaluating the colour heterogeneity has been developed based on dividing the regions of colour according to the perception threshold of the human eye. The method adjusts each point cloud into an ellipsoid within the CIELAB colour space, and its orientation, size, and position depend on the chromatic characteristics of the sample. This method improves the way of expressing and quantifying the chromatic heterogeneity of food products.

FOURTH. The position, dimensions and orientation of the ellipsoids obtained from the point clouds have shown efficient for predicting chemical characteristics of white grape seeds. These parameters were successfully applied for classifying the seeds within a sample into different stages of ripening. This method allowed finding patterns of evolution during the ripening for each variety of grape seeds studied.

FIFTH. The appearance characteristics obtained by image analysis were quite useful for evaluating the chemical composition of grape seeds. This way, a

---

high correlation between the phenolic ripening of seeds and their colorimetric coordinates was found by applying multivariate statistic techniques to chemical data and image analysis results.

SIXTH. The effectiveness of hyperspectral imaging for evaluating the grape seeds has been proved. From the reflectance spectra of grape seeds it was possible to discriminate the seeds according to grape variety, and even according to agronomical factors such as the kind of soil,

SEVENTH. The effectiveness of the reduction of variables by principal component analysis has been proved for the varietal discrimination. The results obtained from the selected bands were comparable to those obtained from the whole near infrared spectrum. Moreover, the selected bands coincided with those wavelengths that gave the most relevant chemical information in previous studies.

EIGHTH. It is possible to estimate the flavanolic content of grape seeds from the near infrared reflectance spectra by applying the suitable chemometric tools. The ability of prediction improves when the model considers only the flavanols that passes from seeds to the must during the winemaking process, instead of the total flavanol content.

NINTH. The imaging techniques developed, both in visible and in infrared range, are not yet a substitute for conventional chemical analysis, but they arise as a good alternative for estimating the chemical characteristics of grape seeds due to their simplicity, versatility, quickness and low-cost. Moreover, they do not need sample preparation or reagents, thus they may be potentially useful for the rapid assessment of the stage of ripening of grape seeds for the winemaking industry.



# ***REFERENCIAS***





---

## REFERENCIAS

Abbott, J.A. (1999). Quality measurement of fruits and vegetables. *Postharvest Biology and Technology* **15** (3), 207-225.

Adams, E.Q. (1923). A Theory of Color Vision. *Psychological Review* **30** (1), 56-76.

Arzate-Vázquez, I.; Chanona-Pérez, J.J.; de Perea-Flores, M.J.; Calderón-Domínguez, G.; Moreno-Armendérez, M.A.; Calvo, H.; Godoy-Calderón, S.; Quevedo, R.; Gutiérrez-López, G. (2011). Image Processing Applied to Classification of Avocado Variety Hass (*Persea americana* Mill.) During the Ripening Process. *Food and Bioprocess Technology* **4** (7), 1307-1313.

Austin, R. B. (1972). Effects of Environment Before Harvesting on Viability. In: *Viability of Seeds*. Springer Netherlands. pp. 114-149.

Baiano, A.; Terracone, C.; Peri, G.; Romaniello, R. (2012). Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes. *Computers and Electronics in Agriculture* **87** (0), 142-151.

Barbin, D.F.; ElMasry, G.; Sun, D.W.; Allen, P. (2013). Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging. *Food Chemistry* **138** (2-3), 1162-1171.

Bautista-Ortín, A.B.; Rodríguez-Rodríguez, P.; Gil-Muñoz, R.; Juménez-Pascual, E.; Busse-Valverde, N.; Martínez-Cutillas, A.; López-Roca, J.M.; Gómez-Plaza, E. (2012). Influence of berry ripeness on concentration, qualitative composition and extractability of grape seed tannins. *Australian Journal of Grape and Wine Research* **18** (2), 123-130.

Baye, T.M.; Pearson, T.C.; Settles, A.M. (2006). Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *Journal of Cereal Science* **43** (2), 236-243.

Bergqvist, J.; Dokoozlian, N.; Ebisuda, N. (2001). Sunlight Exposure and Temperature Effects on Berry Growth and Composition of Cabernet Sauvignon and Grenache in the Central San Joaquin Valley of California. *American Journal of Enology and Viticulture* **52** (1), 1-7.

Berns, R. S. (2000). *Billmeyer and Saltzman's principles of color technology*. Wiley. New York.

## REFERENCIAS

---

Bokobza, L. (2007). Origin of Near-Infrared Absorption Bands. In: Near-Infrared Spectroscopy: Principles, Instruments, Applications. Wiley-VCH Verlag GmbH. Weinheim, Germany. pp. 11-41.

Boulton, R. (2001). The Copigmentation of Anthocyanins and Its Role in the Color of Red Wine: A Critical Review. *American Journal of Enology and Viticulture* **52** (2), 67-87.

Brereton, R. G. (1992). Multivariate Pattern Recognition in Chemometrics: Illustrated by Case Studies. Elsevier Science.

Brosnan, T.; Sun, D.W. (2004). Improving quality inspection of food products by computer vision--a review. *Journal of Food Engineering* **61** (1), 3-16.

Bruno, T. J. and Svoronos, P. D. N. (2005). CRC Handbook of Fundamental Spectroscopic Correlation Charts. Taylor & Francis.

Unscrambler version 9.7. CAMO, (2007). Trondheim, Norway.

Cheyrier, V.; Rigaud, J. (1986). HPLC Separation and Characterization of Flavonols in the Skins of *Vitis vinifera* var. Cinsault. *American Journal of Enology and Viticulture* **37** (4), 248-252.

Choudhary, R.; Mahesh, S.; Paliwal, J.; Jayas, D.S. (2009). Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosystems Engineering* **102** (2), 115-127.

CIE (1932). Commission Internationale de l'Éclairage, Proceedings 1931. Cambridge University Press, 19.

CIE (2007). Commission internationale de l'Eclairage. Standard Illuminants for Colorimetry. *ISO 11664-2:2007*.

Cubero, S.; Aleixos, N.; Moltó, E.; Gómez-Sanchis, J.; Blasco, J. (2011). Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables. *Food Bioprocess Technol* **4** (4), 487-504.

De Freitas, V.A.P.; Glories, Y. (1999). Concentration and compositional changes of procyanidins in grape seeds and skin of white *Vitis vinifera* varieties. *Journal of the Science of Food and Agriculture* **79** (12), 1601-1606.

Delgado, C. (1995). Un poco de historia. In: El libro del vino. Ediciones del Prado. pp. 10-19.

---

Du, C. J. and Sun, D. W. (2008). Quality Evaluation of Pizzas. In: Computer Vision Technology for Food Quality Evaluation. Academic Press. Amsterdam. pp. 427-446.

Escribano-Bailón, M.T.; Alvarez-Garcia, M.; Rivas-Gonzalo, J.C.; Heredia, F.J.; Santos-Buelga, C. (2001). Color and Stability of Pigments Derived from the Acetaldehyde-Mediated Condensation between Malvidin 3-O-Glucoside and (+)-Catechin. *Journal of Agricultural and Food Chemistry* **49** (3), 1213-1217.

Fernández-Vázquez, R.; Stinco, C.; Meléndez-Martínez, A.J.; Heredia, F.J.; Vicario, I.M. (2011). Visual and instrumental evaluation of orange juice color: A consumers' preference study. *Journal of Sensory Studies* **26** (6), 436-444.

Ferrer-Gallego, R.; García-Marino, M.; Hernández-Hierro, J.M.; Rivas-Gonzalo, J.C.; Escribano-Bailón, M.T. (2010). Statistical correlation between flavanolic composition, colour and sensorial parameters in grape seed during ripening. *Analytica Chimica Acta* **660** (1-2), 22-28.

Ferrer-Gallego, R.; Hernández-Hierro, J.M.; Rivas-Gonzalo, J.C.; Escribano-Bailón, M.T. (2011). Determination of phenolic compounds of grape skins during ripening by NIR spectroscopy. *LWT - Food Science and Technology* **44** (4), 847-853.

Finlayson, G.; Xu, R. (2003). Illuminant and gamma comprehensive normalisation in logRGB space. *Pattern Recognition Letters* **24** (11), 1679-1690.

Foca, G.; Masino, F.; Antonelli, A.; Ulrici, A. (2011). Prediction of compositional and sensory characteristics using RGB digital images and multivariate calibration techniques. *Analytica Chimica Acta* **706** (2), 238-245.

Foster, R. (1969). Organic charge-transfer complexes. Academic Press. New York.

García de Luján, A., Puertas, G. B., and Lara, B. M. (1990). Variedades de vid en Andalucía. Dirección General de Investigación y Extensión Agrarias. Sevilla.

García-Marino, M.; Rivas-Gonzalo, J.C.; Ibáñez, E.; García-Moreno, C. (2006). Recovery of catechins and proanthocyanidins from winery by-products using subcritical water extraction. *Analytica Chimica Acta* **563** (1-2), 44-50.

Gat, N. (1998). Real-time multi- and hyper-spectral imaging for remote sensing and machine vision: an overview.

Gawel, R. (1998). Red wine astringency: a review. *Australian Journal of Grape and Wine Research* **4** (2), 74-95.

## REFERENCIAS

---

Geladi, P.; MacDougall, D.; Martens, H. (1985). Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy* **39** (3), 491-500.

Geladi, P.; Isaksson, H.; Lindqvist, L.; Wold, S.; Esbensen, K. (1989). Principal component analysis of multivariate images. *Chemometrics and Intelligent Laboratory Systems* **5** (3), 209-220.

Glasser, L.G.; McKinney, A.H.; Reilly, C.D.; Schnelle, P.D. (1958). Cube-Root Color Coordinate System. *Journal of the Optical Society of America* **48** (10), 736-740.

Glasser, L.G.; Troy, D.J. (1952). A New High Sensitivity Differential Colorimeter. *Journal of the Optical Society of America* **42** (9), 652-657.

Gómez-Míguez, M.; González-Miret, M.L.; Heredia, F.J. (2007). Evolution of colour and anthocyanin composition of Syrah wines elaborated with pre-fermentative cold maceration. *Journal of Food Engineering* **79** (1), 271-278.

González Marcos, A., Martínez de Pisón Ascacibar, F. J., Pernía Espinoza, A. V., Alba Elías, F., Castejón Limas, M., Ordieres Meré, J., and Vergara González, E. (2006). Técnicas y algoritmos básicos de visión artificial.

González-Miret, M.L.; Ji, W.; Luo, R.; Hutchings, J.; Heredia, F.J. (2007). Measuring colour appearance of red wines. *Food Quality and Preference* **18** (6), 862-871.

Goodchild, A.V.; El Haramain, F.J.; Abd El Moneim, A.; Makkar, H.P.S.; Williams, P.C. (1998). Prediction of phenolics and tannins in forage legumes by near infrared reflectance. *Journal of Near Infrared Spectroscopy* **6** (1-4), 175-181.

Gordillo, B.; Rodríguez-Pulido, F.J.; Escudero-Gilete, M.L.; González-Miret, M.L.; Heredia, F.J. (2012). Comprehensive colorimetric study of anthocyanic copigmentation in model solutions. Effects of pH and molar ratio. *Journal of Agricultural and Food Chemistry* **60** (11), 2896-2905.

Gowen, A.A.; O'Donnell, C.P.; Cullen, P.J.; Downey, G.; Frias, J.M. (2007). Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology* **18** (12), 590-598.

Heredia, F.J. et al. (2006). DigiFood.

Hernández-Hierro, J.M.; Valverde, J.; Villacreces, S.; Reilly, K.; Gaffney, M.; González-Miret, M.L.; Heredia, F.J.; Downey, G. (2012). Feasibility study on the use of visible-near-infrared spectroscopy for the screening of individual and total glucosinolate

---

contents in broccoli. *Journal of Agricultural and Food Chemistry* **60** (30), 7352-7358.

Hernández-Hierro, J.M. (2008). Tesis doctoral: Utilización de la tecnología NIRS con sonda de fibra óptica para el control de calidad de alfalfa.

Hung, P.C. (1991). Colorimetric calibration for scanners and media. *Proceedings of SPIE* **1448**, 164-174.

Hung, P.C. (1993). Colorimetric calibration in electronic imaging devices using a look-up-table model and interpolations. *Journal of Electronic Imaging* **2** (1), 53-61.

Hutchings, J. (1999). *Food Color and Appearance*. Springer.

Hutchings, J., Luo, R., and Ji, W. (2002). Calibrated colour imaging analysis of food. In: *Colour in food: improving quality*. CRC Press.

International Organization for Standardization (1999). *Graphic technology and photography: Colour characterisation of digital still cameras (DSCs) Part 1: Stimuli, metrology, and test procedures (ISO 17321-1:1999)*.

Jackman, P.; Sun, D.W.; Du, C.J.; Allen, P. (2009). Prediction of beef eating qualities from colour, marbling and wavelet surface texture features using homogenous carcass treatment. *Pattern Recognition* **42** (5), 751-763.

Jackson, J. E. (2005). *A User's Guide to Principal Components*. Wiley.

Jaglarlz, J.; Duraj, R.; Szopa, P.; Cisowski, J.; Czternastek, H. (2006). Investigation of white standards by means of bidirectional reflection distribution function and integrating sphere methods. *Optica Applicata* **36** (1), 97-103.

Johnson, D. E. (2000). *Métodos multivariados aplicados al análisis de datos*. International Thomson Editores, S. A. Madrid.

Judd, D.B. (1949). Response functions for types of vision according to the Müller theory. *Journal of Research of the National Bureau of Standards* **42** (1), 1-15.

Judd, D. B. and Wyszecki, G. (1975). *Color in business, science, and industry*. Wiley.

Kang, H.R. (1992). Colour scanner calibration. *Journal of Imaging Science and Technology* **36** (2 (March/April)).

## REFERENCIAS

---

Kang, H.R.; Anderson, P.G. (1992). Neural network applications to the color scanner and printer calibrations. *Journal of Electronic Imaging* **1** (2), 125-135.

Kennedy, J.A.; Matthews, M.A.; Waterhouse, A.L. (2000a). Changes in grape seed polyphenols during fruit ripening. *Phytochemistry* **55** (1), 77-85.

Kennedy, J.A.; Troup, G.J.; Pilbrow, J.R.; HUTTON, D.R.; Hewitt, D.; HUNTER, C.R.; Ristic, R.; Iland, P.G.; JONES, G.P. (2000b). Development of seed polyphenols in berries from *Vitis vinifera* L. cv. Shiraz. *Australian Journal of Grape and Wine Research* **6** (3), 244-254.

Koehler IV, F.W.; Lee, E.; Kidder, L.H.; Lewis, N.E. (2002). Near infrared spectroscopy: the practical chemical imaging solution. *Spectroscopy Europe* **14**, 12-19.

Lacar, F.M.; Lewis, M.M.; Grierson, I.T. (2001). Use of hyperspectral reflectance for discrimination between grape varieties. *Geoscience and Remote Sensing Symposium, 2001. IGARSS '01. IEEE 2001 International* **6**, 2878-2880.

León, K.; Mery, D.; Pedreschi, F.; León, J. (2006). Color measurement in L\*a\*b\* units from RGB digital images. *Food Research International* **39** (10), 1084-1091.

Li, C.; Wang, B.; Wang, J.; Li, F. (2012). Extracting vein of leaf image based on K-means clustering. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering* **28** (17), 157-162.

Liu, L.; Cozzolino, D.; Cynkar, W.U.; Gishen, M.; Colby, C.B. (2006). Geographic Classification of Spanish and Australian Tempranillo Red Wines by Visible and Near-Infrared Spectroscopy Combined with Multivariate Analysis. *Journal of Agricultural and Food Chemistry* **54** (18), 6754-6759.

Liu, Y.-X.; Yan, G.-L.; He, J.-J.; Duan, C.-Q. (2010). Changes of Flavan-3-ols with Different Degrees of Polymerization in Seeds of 'Shiraz', 'Cabernet Sauvignon' and 'Marselan' Grapes after Veraison. *Molecules* **15** (11), 7763-7774.

Lorente, D.; Aleixos, N.; Gómez-Sanchis, J.; Cubero, S.; García-Navarrete, O.L.; Blasco, J. (2012). Recent Advances and Applications of Hyperspectral Imaging for Fruit and Vegetable Quality Assessment. *Food Bioprocess Technol* **5** (4), 1121-1142.

Luke, J.T. (1999). OSA Instrumental in Development of the uniform color scales. *Optics and Photonics News* **10** (9), 28-33.

Luo, M.R. et al. (2001). British Patent entitled Apparatus and method for measuring colour (DigiEye System), Derby University Enterprises Limited.



---

MacAdam, D.L. (1942). Visual Sensitivities to Color Differences in Daylight. *Journal of the Optical Society of America* **32** (5), 247-273.

Magnan, P. (2003). Detection of visible photons in CCD and CMOS: A comparative view. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **504** (1ΓÇô3), 199-212.

Martín, Q. (2001). Curso de tratamiento de datos estadísticos con SPSS. Hespérides. Salamanca.

Martínez, J.A.; Melgosa, M.; Pérez, M.M.; Hita, E.; Negueruela, A.I. (2001). Note. Visual and Instrumental Color Evaluation in Red Wines. *Food Science and Technology International* **7** (5), 439-444.

Maxwell, J.C. (1860). On the Theory of Compound Colours, and the Relations of the Colours of the Spectrum. *Philosophical Transactions of the Royal Society* **150**, 57-84.

McCamy, C.S.; Marcus, H.; Davidson, J.G. (1976). A Color-Rendition Chart. *Journal of Applied Photographic Engineering* **2** (3), 95-99.

Mehl, P.M.; Chen, Y.R.; Kim, M.S.; Chan, D.E. (2004). Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering* **61** (1), 67-81.

Melgosa, M.; Pérez, M.M.; Yebra, A.; Huertas, R.; Hita, E. (2001). Algunas reflexiones y recientes recomendaciones internacionales sobre evaluación de diferencias de color. *Óptica Pura y Aplicada* **34** (1), 1-10.

Mendoza, F.; Dejmek, P.; Aguilera, J.M. (2006). Calibrated color measurements of agricultural foods using image analysis. *Postharvest Biology and Technology* **41** (3), 285-295.

Mullins, M. G., Bouquet, A., and Williams, L. E. (1992). The structure of the grapevine: vegetative and reproductive anatomy. In: *Biology of the Grapevine*. Cambridge University Press. pp. 37-79.

Munck, L.; Møller, B. (2004). A new germinative classification model of barley for prediction of malt quality amplified by a near infrared transmission spectroscopy calibration for vigour "on line" both implemented by multivariate data analysis. *Journal of the Institute of Brewing* **110** (1), 3-17.

Nalwa, V. S. (1993). A guided tour of computer vision. Addison-Wesley.

## REFERENCIAS

---

Nansen, C.; Kolomiets, M.; Gao, X. (2008). Considerations Regarding the Use of Hyperspectral Imaging Data in Classifications of Food Products, Exemplified by Analysis of Maize Kernels. *Journal of Agricultural and Food Chemistry* **56** (9), 2933-2938.

Newton, I. (1671). New Theory about Light and Colors. *Philosophical Transactions of the Royal Society* **6** (69-80), 3075-3087.

Nicolai, B.M.; Beullens, K.; Bobelyn, E.; Peirs, A.; Saeys, W.; Theron, K.I.; Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology* **46** (2), 99-118.

Norman, G. R., De Burgh Norman, J. E., and Streiner, D. L. (1996). Bioestadística. Harcourt.

OIV (2013a). Compendium of international methods of wine and must analysis. The International Organisation of Vine and Wine. Paris.

OIV (2013b). Situation et statistiques du secteur vitivinicole mondial. Paris.

Osborne, B. G., Fearn, T., and Hindle, P. T. (1993). Practical NIR Spectroscopy With Applications in Food and Beverage Analysis. Longman Group. United Kingdom.

Pacifici, S. (2012). Análisis de densidades mamográficas en espacio RGB. *Imagen Diagnóstica* **3** (2), 45-49.

Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society* **14** (2), 198-219.

Pauli, H. (1976). Proposed extension of the CIE recommendation on "Uniform color spaces, color difference equations, and metric color terms". *Journal of the Optical Society of America* **66** (8), 866-867.

Peyrot des Gachons, C.; Kennedy, J.A. (2003). Direct method for determining seed and skin proanthocyanidin extraction into red wine. *Journal of Agricultural and Food Chemistry* **51** (20), 5877-5881.

Pratt, C. (1971). Reproductive Anatomy in Cultivated Grapes - A Review. *American Journal of Enology and Viticulture* **22** (2), 92-109.

Real Academia Española (2001). Diccionario de la lengua española (22<sup>a</sup> ed.). Espasa Libros.

---

Ribereau-Gayon, P., Glories, Y., Maujean, A., and Dubourdieu, D. (2006). Phenolic Compounds. In: Handbook of Enology. John Wiley & Sons, Ltd. pp. 141-203.

Ristic, R.; Iland, P.G. (2005). Relationships between seed and berry development of *Vitis Vinifera* L. cv Shiraz: Developmental changes in seed morphology and phenolic composition. *Australian Journal of Grape and Wine Research* **11** (1), 43-58.

Robinson, J. (1996). Jancis Robinson's guide to wine grapes. Oxford University Press.

Rousseau, J.; Delteil, D. (2000). Présentation d'une méthode d'analyse sensorielle des raisins. Principe, méthode et grille d'interprétation. *Revue française d'oenologie* **183**, 10-13.

Sandhu, A.K.; Gu, L. (2010). Antioxidant Capacity, Phenolic Content, and Profiling of Phenolic Compounds in the Seeds, Skin, and Pulp of *Vitis rotundifolia* (Muscadine Grapes) As Determined by HPLC-DAD-ESI-MSn. *Journal of Agricultural and Food Chemistry* **58** (8), 4681-4692.

Santos-Buelga, C.; Francia-Aricha, E.M.; Escribano-Bailón, M.T. (1995). Comparative flavan-3-ol composition of seeds from different grape varieties. *Food Chemistry* **53** (2), 197-201.

Savakar, D.G.; Anami, B.S. (2009). Classification of Food Grains, Fruits and Flowers Using Machine Vision. *International Journal of Food Engineering* **5** (4), 14.

Schultz, H.R.; Jones, G.V. (2010). Climate Induced Historic and Future Changes in Viticulture. *Journal of Wine Research* **21** (2-3), 137-145.

Shahin, M.A.; Symons, S.J. (2011). Detection of Fusarium damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis. *Computers and Electronics in Agriculture* **75** (1), 107-112.

Shaw, G.; Manolakis, D. (2002). Signal processing for hyperspectral image exploitation. *Signal Processing Magazine, IEEE* **19** (1), 12-16.

Statistica 8.0 StatSoft Inc., (2007). Tulsa, USA.

Sugiyama, T.; Sugiyama, J.; Tsuta, M.; Fujita, K.; Shibata, M.; Kokawa, M.; Araki, T.; Nabetani, H.; Sagara, Y. (2010). NIR spectral imaging with discriminant analysis for detecting foreign materials among blueberries. *Journal of Food Engineering* **101** (3), 244-252.

## REFERENCIAS

---

Sun, D.W. (2000). Inspecting pizza topping percentage and distribution by a computer vision method. *Journal of Food Engineering* **44** (4), 245-249.

Sun, D.W. (2004). Computer vision--an objective, rapid and non-contact quality evaluation tool for the food industry. *Journal of Food Engineering* **61** (1), 1-2.

Tabachnick, B. G. and Fidell, L. S. (2012). Using Multivariate Statistics. Pearson Education, Limited.

MATLAB version R2012b. The MathWorks Inc., (2012). Natick, USA.

Thorngate, J.H.; Singleton, V.L. (1994). Localization of Procyanidins on Grape Seeds. *American Journal of Enology and Viticulture* **45** (2), 259-262.

Valous, N.A.; Mendoza, F.; Sun, D.W.; Allen, P. (2009). Colour calibration of a laboratory computer vision system for quality evaluation of pre-sliced hams. *Meat Science* **81** (1), 132-141.

Vivas, N.; Glories, Y.; Lagune, L.; Saucier, C.; Augustin, M. (1994). Estimation du degré de polymérisation des procyanidines du raisin et du vin par la méthode au p-diméthylaminocinnamaldéhyde. *Journal International des Sciences de la Vigne et du Vin* **28**, 319-336.

Wang, H.H.; Sun, D.W. (2002). Correlation between Cheese Meltability Determined with a Computer Vision Method and with Arnott and Schreiber Tests. *Journal of Food Science* **67** (2), 745-749.

Wang, H.H.; Sun, D.W. (2001). Evaluation of the functional properties of Cheddar Cheese using a computer vision method. *Journal of Food Engineering* **49** (1), 49-53.

Waterhouse, A.L. (2002). Wine Phenolics. *Annals of the New York Academy of Sciences* **957**, 21-36.

Werker, E. (1980). Seed dormancy as explained by the anatomy of embryo envelopes. *Israel Journal of Botany* **29** (1-4), 22-44.

Werker, E. (1997). Seed Anatomy. In: Encyclopedia of plant anatomy. Borntraeger. Berlin.

Williams, P. C. (1987). Variables affecting Near-infrared Reflectance Spectroscopic Analysis. In: Near-infrared Technology in the Agriculture and Food Industries. American Association of Cereal Chemists. Minnesota. pp. 143-167.

---

Williams, P.; Geladi, P.; Fox, G.; Manley, M. (2009). Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta* **653** (2), 121-130.

Williams, P. and Norris, K. (2001). Near-Infrared Technology: In the Agricultural and Food Industries. American Association of Cereal Chemists. St Paul, MN.

Wold, J.P.; Jakobsen, T.; Krane, L. (1996). Atlantic Salmon Average Fat Content Estimated by Near-Infrared Transmittance Spectroscopy. *Journal of Food Science* **61** (1), 74-77.

Wyszecki, G. (1954). A Regular Rhombohedral Lattice Sampling of Munsell Renotation Space. *Journal of the Optical Society of America* **44** (9), 725-725.

Wyszecki, G. and Stiles, W. S. (1982). Color Science: Concepts and Methods, Quantitative Data and Formulae. John Wiley & Sons. New York.

Xia, E.-Q.; Deng, G.-F.; Guo, Y.-J.; Li, H.-B. (2010). Biological Activities of Polyphenols from Grapes. *International Journal of Molecular Sciences* **11** (2), 622-646.

Xing, J.; Symons, S.; Shahin, M.; Hatcher, D. (2010). Detection of sprout damage in Canada Western Red Spring wheat with multiple wavebands using visible/near-infrared hyperspectral imaging. *Biosystems Engineering* **106** (2), 188-194.

Xing, J.; Van Hung, P.; Symons, S.; Shahin, M.; Hatcher, D. (2009). Using a Short Wavelength Infrared (SWIR) hyperspectral imaging system to predict alpha amylase activity in individual Canadian western wheat kernels. *Sensing and Instrumentation for Food Quality and Safety* **3** (4), 211-218.

Yager, R.R.; Filev, D.P. (1994). Generation of Fuzzy Rules by Mountain Clustering. *Journal of Intelligent and Fuzzy Systems* **2**, 209-219.

Yam, K.L.; Papadakis, S.E. (2004). A simple digital imaging method for measuring and analyzing color of food surfaces. *Journal of Food Engineering* **61** (1), 137-142.

Young, T. (1802a). An Account of Some Cases of the Production of Colours, not Hitherto Described. *Philosophical Transactions of the Royal Society* **92**, 387-397.

Young, T. (1802b). On the Theory of Light and Colours. *Philosophical Transactions of the Royal Society* **92**, 12-48.

Zamora-Marín, F. (2004). El "délestaje": una técnica muy útil para la elaboración de vinos tintos. *Enólogos* **37**, 28-31.

## REFERENCIAS

---

Zamora-Marín, F. (2006). El cambio climático, una amenaza para nuestra viticultura. *Enólogos* **29**, 28-31.

Zheng, C. and Sun, D. W. (2008). Object Measurement Methods. In: *Computer Vision Technology for Food Quality Evaluation*. Academic Press. Amsterdam. pp. 57-80.

Zoecklein, B. W., Fugelsang, K., Gump, B., and Nury, F. (2001). *Análisis y producción de vino*. Ed. Acribia. Zaragoza.

***ANEXO***







---

## SOFTWARE DESARROLLADO PARA ESTA TESIS (LENGUAJE MATLAB)

NOTA: El símbolo “▶” significa que en el programa original no existe salto de línea.

### 1. Software para el análisis morfológico y colorimétrico de las semillas de uva.

Utilizado en los Capítulos 1, 2 y 3.

**DigiSeed** – *Registro territorial de la Propiedad Intelectual de Andalucía*

*(Expediente SE-471-12)*

```
function colorsemilla=DigiSeed(directorio)
matrizcambio=[0.001476931 3.009220943;
0.003954236 -129.5569476;
0.00393619 -129.0494204];
cd(directorio);
listaarchivos(:,1)=dir('*RGB.TIF');
listaarchivos(:,2)=dir('*LAB.TIF');
tamanolista=size(listaarchivos);
for contadorprimero=1:tamanolista(1,1)
semillainicialLAB=imread(listaarchivos(contadorprimero,2).name);
semillainicial=imread(listaarchivos(contadorprimero,1).name);
tamano=size(semilla);
semilla=imcrop(semillainicial,[1 1 round(tamano(1,2)/2) tamano(1,1)]);
semillaLAB=imcrop(semillainicialLAB,[1 1 round(tamano(1,2)/2)
tamano(1,1)]);
clear semillainicial semillainicialLAB
tamano=size(semilla);
mascarahuecos=zeros(tamano(1,1),tamano(1,2));
for i=(1:tamano(1,1))
for j=(1:tamano(1,2))
if (semilla(i,j,1)>45) && (semilla(i,j,1)<190) && (semilla(i,j,2)>45)▶
&&(semilla(i,j,2)<195) && (semilla(i,j,3)>45) && (semilla(i,j,3)<160)▶
mascarahuecos(i,j)=1;
end
end
end
mascara=(imfill(mascarahuecos));
clear mascarahuecos;
discoerosion=strel('disk',7);
mascaraerosion=imerode(mascara,discoerosion);
clear mascara;
clear discoerosion;
discodilatacion=strel('disk',3);
mascarafinal=imdilate(mascaraerosion,discodilatacion);
clear mascaraerosion;
clear discodilatacion;
semillaCIELAB=double(zeros(tamano(1),tamano(2),3));
for i=1:tamano(1)
for j=1:tamano(2)
for k=1:3
semillaCIELAB(i,j,k)=(double(semillaLAB(i,j,k))*matrizcambio(k,1))▶
+matrizcambio(k,2);
end
end
end
```

---

```

clear semillaLAB
mascaradefinitiva=double(mascarafinal);
imagen(:,:,1)=double(mascaradefinitiva.*semillaCIELAB(:,:,1));
imagen(:,:,2)=double(mascaradefinitiva.*semillaCIELAB(:,:,2));
imagen(:,:,3)=double(mascaradefinitiva.*semillaCIELAB(:,:,3));
tablaLAB=reshape(imagen,tamano(1,1)*tamano(1,2),3);
tablaordenada=sortrows(tablaLAB,-1);
clear tablaLAB
tamanotablaLAB=size(tablaordenada);
i=1;
while i<tamanotablaLAB(1,1)
if tablaordenada(i,1)==0
tablaordenada(i:tamanotablaLAB(1,1),:)=[];
i=tamanotablaLAB(1,1);
end
i=i+1;
end
tamano=size(tablaordenada);
for i=1:tamano(1,1);
tablaordenada(i,4)=sqrt((tablaordenada(i,2))^2+▶
(tablaordenada(i,3)^2));
if ((tablaordenada(i,3)>0)&&(tablaordenada(i,2)<0)) || ▶
((tablaordenada(i,3)<0)&&(tablaordenada(i,2)<0))tablaordenada(i,5)= ▶
(180/pi())*atan(tablaordenada(i,3)/tablaordenada(i,2))+180;
if (tablaordenada(i,5)<0)
tablaordenada(i,5)=360+tablaordenada(i,5);
end
else
tablaordenada(i,5)=(180/pi())*atan(tablaordenada(i,3)/▶
tablaordenada(i,2));
if (tablaordenada(i,5)<0)
tablaordenada(i,5)=360+tablaordenada(i,5);
end
end
end
paso=ceil((tamano(1,1))/500);
colorsemilla=tablaordenada(1:paso:tamano(1,1),:);
clear semilla tablaordenada tamano tamanotablaLAB i j paso ans
nombre=strcat(listaarchivos(contadorprimero,1).name, '_color.txt');
save(nombre, 'colorsemilla', '-ASCII');
media=[mean(colorsemilla(:,1)) std(colorsemilla(:,1))
mean(colorsemilla(:,2)) std(colorsemilla(:,2)) ▶
mean(colorsemilla(:,3)) std(colorsemilla(:,3)) ▶
mean(colorsemilla(:,4)) std(colorsemilla(:,4)) ▶
mean(colorsemilla(:,5)) std(colorsemilla(:,5))];
nombre=strcat(listaarchivos(contadorprimero,1).name,▶
'_mediacolor.txt');
save(nombre, 'media', '-ASCII');
clear nombre colorsemilla
end
listatxt=dir('*.TIF_color.txt');
tablafinal=[0 0 0 0 0];
nombres=[listatxt(1,1).name];
for i=1:size(listatxt,1)
variabletemporal=load(listatxt(i,1).name);
for j=1:size(variabletemporal,1)
nombres=[nombres;listatxt(i,1).name];
end
tablafinal=[tablafinal; variabletemporal];

```

---

---

```

end
tablafinal(1,:)=[];nombres(1,:)=[];
nombres=cellstr(nombres);
xlswrite('Color.xls',nombres,'datos','A2');
xlswrite('Color.xls',tablafinal,'datos','B2');
xlswrite('Color.xls',{'L*','a*','b*','C*ab','hab'},'datos','B1');
listamedias=dir('*.TIF_mediacolor.txt');
tablafinalmedias=[0 0 0 0 0 0 0 0 0 0];
nombres=[listamedias(1,1).name];
for i=1:size(listamedias,1)
variabletemp=load(listamedias(i,1).name);
for j=1:size(variabletemp,1)
nombres=[nombres;listamedias(i,1).name];
end
tablafinalmedias=[tablafinalmedias; variabletemp];
end
tablafinalmedias(1,:)=[];nombres(1,:)=[];
nombres=cellstr(nombres);
xlswrite('MediasColor.xls',nombres,'datos','A2');
xlswrite('MediasColor.xls',tablafinalmedias,'datos','B2');
xlswrite('MediasColor.xls',{'L*','desv_L*','a*','desv_a*','b*','▶
'desv_b*','C*ab','desv_C*ab','hab','desv_hab'},'datos','B1');
recycle('*color.txt');
disp('Color de todos los archivos calculado.')
for contadorprimero=1:tamanolista(1,1)
semilla=imread(listaarchivos(contadorprimero,1).name);
tamano=size(semilla);
mascarahuecos=zeros(tamano(1,1),tamano(1,2));
mascarahuecos=uint8(mascarahuecos);
for i=(1:tamano(1,1))
for j=(1:tamano(1,2))
if(semilla(i,j,1)>45) && (semilla(i,j,1)<190) && (semilla(i,j,2)>45)▶
&&(semilla(i,j,2)<195) && (semilla(i,j,3)>45) && ▶
(semilla(i,j,3)<160)
mascarahuecos(i,j)=1;
end
end
end
mascara=(imfill(mascarahuecos));
clear mascarahuecos;
for i=((round((tamano(1,1))/2)-100):(round((tamano(1,1))/2)+100));
for j=((round((tamano(1,2))/2)-100):(round((tamano(1,2))/2)+100));
mascara(i,j)=0;
end
end
discoerosion=strel('disk',7);
mascaraerosion=imerode(mascara,discoerosion);
clear mascara;
clear discoerosion;
discodilatacion=strel('disk',7);
mascarafinal=imdilate(mascaraerosion,discodilatacion);
clear mascaraerosion;
clear discodilatacion;
mascaraetiquetada=bwlabel(mascarafinal);
for i=(1:tamano(1,1))
for j=(round((tamano(1,2))/2):tamano(1,2))
mascaraetiquetada(i,j)=0;
end
end

```

---

---

```

objetos=max(max(mascaraetiquetada));
clear mascarafinal;
morfologia=regionprops
(mascaraetiquetada,'MajorAxisLength','MinorAxisLength','Area',►
'eccentricity','EquivDiameter','Perimeter');
for i=(1:objetos)
lista(i,1)=morfologia(i,1).MajorAxisLength;
lista(i,2)=morfologia(i,1).MinorAxisLength;
lista(i,3)=morfologia(i,1).Area;
lista(i,4)=morfologia(i,1).Eccentricity;
lista(i,5)=morfologia(i,1).EquivDiameter;
lista(i,6)=morfologia(i,1).Perimeter;
lista(i,8)=objetos;
end
lista(:,7)=((lista(:,6)).^2)./(4*pi().*(lista(:,3)));
lista(:,1)=lista(:,1)/11.061465001;
lista(:,2)=lista(:,2)/11.061465001;
lista(:,5)=lista(:,5)/11.061465001;
lista(:,6)=lista(:,6)/11.061465001;
lista(:,3)=lista(:,3)/((11.061465001)^2);
clear i j objetos tamano morfologia semilla mascaraetiquetada
nombre=strcat(listaarchivos(contadorprimero,1).name,'_morf.txt');
save(nombre,'lista','-ASCII');
media=[mean(lista(:,1)) std(lista(:,1)) mean(lista(:,2)) ►
std(lista(:,2)) mean(lista(:,3)) std(lista(:,3)) mean(lista(:,4)) ►
std(lista(:,4)) mean(lista(:,5)) std(lista(:,5)) mean(lista(:,6)) ►
std(lista(:,6)) mean(lista(:,7)) std(lista(:,7)) lista(1,8)];
nombre=strcat(listaarchivos(contadorprimero,1).name,'_mediamorf.txt');
save(nombre,'media','-ASCII');
clear nombre lista
end
listatxt=dir('*.TIF_morf.txt');
tablafinal=[0 0 0 0 0 0 0 0];
nombres=[listatxt(1,1).name];
for i=1:size(listatxt,1)
variabletemporal=load(listatxt(i,1).name);
for j=1:size(variabletemporal,1)
nombres=[nombres;listatxt(i,1).name];
end
end
tablafinal=[tablafinal; variabletemporal];
end
tablafinal(1,:)=[];nombres(1,:)=[];
nombres=cellstr(nombres);
xlswrite('morfologia.xls',nombres,'datos','A2');
xlswrite('morfologia.xls',tablafinal,'datos','B2');
xlswrite('morfologia.xls',{'MajorAxis','MinorAxis','Area',►
'Eccentricity','EquivDiameter','Perimeter','Roundness', ►
'N° Semillas'},'datos','B1');
listamedias=dir('*.TIF_mediamorf.txt');
tablafinalmedias=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
nombres=[listamedias(1,1).name];
for i=1:size(listamedias,1)
variabletemp=load(listamedias(i,1).name);
for j=1:size(variabletemp,1)
nombres=[nombres;listamedias(i,1).name];
end
end
tablafinalmedias=[tablafinalmedias; variabletemp];
end
tablafinalmedias(1,:)=[];nombres(1,:)=[];

```

---

---

```
nombres=cellstr(nombres);
xlswrite('MediasMorfología.xls',nombres,'datos','A2');
xlswrite('MediasMorfología.xls',tablafinalmedias,'datos','B2');
xlswrite('MediasMorfología.xls',{ 'MajorAxis','desv_MajorAxis',▶
'MinorAxis','desv_MinorAxis','Area','desv_Area','Eccentricity',▶
'desv_Eccentricity','EquivDiameter','desv_EquivDiameter',▶
'Perimeter','desv_Perimeter','Roundness','desv_Roundness',▶
'Nº Semillas'},'datos','B1');
recycle('*morf.txt');
disp('Morfología de todos los archivos calculada')
end
```

## 2. Software para el análisis morfológico y colorimétrico de uvas.

Utilizado en los Capítulos 1 y 2.

**DigiGrape** – *Registro territorial de la Propiedad Intelectual de Andalucía*

*(Expediente SE-472-12)*

```
function tablafinal=DigiGrape(directorio)
matrizcambio=[0.001476931 3.009220943;
0.003954236 -129.5569476;
0.00393619 -129.0494204];
cd(directorio);
listaarchivos(:,1)=dir('*RGB.TIF');
listaarchivos(:,2)=dir('*LAB.TIF');
tamanolista=size(listaarchivos);
for contadorprimero=1:tamanolista(1,1)
uvainicialLAB=imread(listaarchivos(contadorprimero,2).name);
uvainicial=imread(listaarchivos(contadorprimero,1).name);
tamaño=size(uvainicial);
uva=imcrop(uvainicial,[1 1 round(tamaño(1,2)/2) tamaño(1,1)]);
uvaLAB=imcrop(uvainicialLAB,[1 1 round(tamaño(1,2)/2) tamaño(1,1)]);
clear uvainicial uvainicialLAB
tamaño=size(uva);
mascarahuecos=zeros(tamaño(1,1),tamaño(1,2));
for i=(1:tamaño(1,1))
for j=(1:tamaño(1,2))
if uva(i,j,3)<135
mascarahuecos(i,j)=1;
end
end
end
mascara=(imfill(mascarahuecos));
clear mascarahuecos;
discoerosion=strel('disk',7);
mascaraerosion=imerode(mascara,discoerosion);
clear mascara;
clear discoerosion;
discodilatacion=strel('disk',7);
mascarafinal=imdilate(mascaraerosion,discodilatacion);
clear mascaraerosion;
clear discodilatacion;
mascaraetiquetada=bwlabel(mascarafinal);
objetos=max(max(mascaraetiquetada(:,:)));
morfologia=regionprops
(mascaraetiquetada,'MajorAxisLength','MinorAxisLength','Area',►
'eccentricity','EquivDiameter','Perimeter');
for i=(1:objetos)
lista(i,1)=morfologia(i,1).MajorAxisLength;
lista(i,2)=morfologia(i,1).MinorAxisLength;
lista(i,3)=morfologia(i,1).Area;
lista(i,4)=morfologia(i,1).Eccentricity;
lista(i,5)=morfologia(i,1).EquivDiameter;
lista(i,6)=morfologia(i,1).Perimeter;
lista(i,8)=objetos;
end
lista(:,7)=((lista(:,6)).^2)./(4*pi().*(lista(:,3)));
lista(:,1)=lista(:,1)/11.061465001;
lista(:,2)=lista(:,2)/11.061465001;
```

---

```

lista(:,5)=lista(:,5)/11.061465001;
lista(:,6)=lista(:,6)/11.061465001;
lista(:,3)=lista(:,3)/((11.061465001)^2);
clear i j objetos morfologia semilla mascaraetiquetada
nombre=strcat(listaarchivos(contadorprimero,1).name, '_morf.txt');
save(nombre, 'lista', '-ASCII');
media=[mean(lista(:,1)) std(lista(:,1)) mean(lista(:,2))▶
std(lista(:,2)) mean(lista(:,3)) std(lista(:,3)) mean(lista(:,4))▶
std(lista(:,4)) mean(lista(:,5)) std(lista(:,5)) mean(lista(:,6))▶
std(lista(:,6)) mean(lista(:,7)) std(lista(:,7)) lista(1,8)];
nombre=strcat(listaarchivos(contadorprimero,1).name, '_mediamorf.txt');
save(nombre, 'media', '-ASCII');
clear nombre lista
disp(strcat(listaarchivos(contadorprimero,1).name, ' calculado ▶
MORFOLOGÍA' ))
uvaCIELAB=double(zeros(tamano(1),tamano(2),3));
for i=1:tamano(1)
for j=1:tamano(2)
for k=1:3
uvaCIELAB(i,j,k)=(double(uvaLAB(i,j,k))*matrizcambio(k,1))▶
+matrizcambio(k,2);
end
end
end
clear uvaLAB
mascaradefinitiva=double(mascarafinal);
clear mascarafinal
imagen(:, :, 1)=double(mascaradefinitiva.*uvaCIELAB(:, :, 1));
imagen(:, :, 2)=double(mascaradefinitiva.*uvaCIELAB(:, :, 2));
imagen(:, :, 3)=double(mascaradefinitiva.*uvaCIELAB(:, :, 3));
tablaLAB=reshape(imagen,tamano(1,1)*tamano(1,2),3);
tablaordenada=sortrows(tablaLAB,-1);
clear tablaLAB
tamanotablaLAB=size(tablaordenada);
i=1;
while i<tamanotablaLAB(1,1)
if tablaordenada(i,1)==0
tablaordenada(i:tamanotablaLAB(1,1),:)=[];
i=tamanotablaLAB(1,1);
end
i=i+1;
end
tamaño=size(tablaordenada);
for i=1:tamaño(1,1);
tablaordenada(i,4)=sqrt((tablaordenada(i,2))^2+(tablaordenada(i,3))^2)▶
);
if ((tablaordenada(i,3)>0)&&(tablaordenada(i,2)<0)) || ▶
((tablaordenada(i,3) <0)&&(tablaordenada(i,2)<0))
tablaordenada(i,5)=(180/pi())*atan(tablaordenada(i,3)/▶
tablaordenada(i,2))+180;
if (tablaordenada(i,5)<0)
tablaordenada(i,5)=360+tablaordenada(i,5);
end
else
tablaordenada(i,5)=(180/pi())*atan(tablaordenada(i,3)/ ▶
tablaordenada(i, 2));
if (tablaordenada(i,5)<0)
tablaordenada(i,5)=360+tablaordenada(i,5);
end
end

```

---

```
end
end
paso=ceil((tamano(1,1))/500);
coloruva=tablaordenada(1:paso:tamano(1,1),:);
clear tablaordenada tamano tamanotablaLAB i j paso ans
nombre=strcat(listaarchivos(contadorprimero,1).name, '_color.txt');
save(nombre, 'coloruva', '-ASCII');
media=[mean(coloruva(:,1)) std(coloruva(:,1)) mean(coloruva(:,2)) ▶
std(coloruva(:,2)) mean(coloruva(:,3)) std(coloruva(:,3)) ▶
mean(coloruva(:,4)) std(coloruva(:,4)) mean(coloruva(:,5)) ▶
std(coloruva(:,5))];
nombre=strcat(listaarchivos(contadorprimero,1).name, ▶
'_mediacolor.txt');
save(nombre, 'media', '-ASCII');
clear nombre coloruva
disp(strcat(listaarchivos(contadorprimero,1).name, ' calculado. ▶
COLOR'))
end
listatxt=dir('*TIF_color.txt');
tablafinal=[0 0 0 0 0];
nombres=[listatxt(1,1).name];
for i=1:size(listatxt,1)
variabletemporal=load(listatxt(i,1).name);
for j=1:size(variabletemporal,1)
nombres=[nombres;listatxt(i,1).name];
end
tablafinal=[tablafinal; variabletemporal];
end
tablafinal(1,:)=[];nombres(1,:)=[];
nombres=cellstr(nombres);
xlswrite('Color.xls',nombres,'datos','A2');
xlswrite('Color.xls',tablafinal,'datos','B2');
xlswrite('Color.xls',{'L*','a*','b*','C*ab','hab'},'datos','B1');
listamedias=dir('*TIF_mediacolor.txt');
tablafinalmedias=[0 0 0 0 0 0 0 0 0 0];
nombres=[listamedias(1,1).name];
for i=1:size(listamedias,1)
variabletemp=load(listamedias(i,1).name);
for j=1:size(variabletemp,1)
nombres=[nombres;listamedias(i,1).name];
end
tablafinalmedias=[tablafinalmedias; variabletemp];
end
tablafinalmedias(1,:)=[];nombres(1,:)=[];
nombres=cellstr(nombres);
xlswrite('MediasColor.xls',nombres,'datos','A2');
xlswrite('MediasColor.xls',tablafinalmedias,'datos','B2');
xlswrite('MediasColor.xls',
{'L*','desv_L*','a*','desv_a*','b*','desv_b*','C*ab','desv_C*ab',▶
'hab','desv_hab'},'datos','B1');
recycle('*color.txt');
listatxt=dir('*TIF_morf.txt');
tablafinal=[0 0 0 0 0 0 0 0];
nombres=[listatxt(1,1).name];
for i=1:size(listatxt,1)
variabletemporal=load(listatxt(i,1).name);
for j=1:size(variabletemporal,1)
nombres=[nombres;listatxt(i,1).name];
end
```



---

```

tablafinal=[tablafinal; variabletemporal];
end
tablafinal(1,:)=[];nombres(1,:)=[];
nombres=cellstr(nombres);
xlswrite('morfologia.xls',nombres,'datos','A2');
xlswrite('morfologia.xls',tablafinal,'datos','B2');
xlswrite('morfologia.xls',
{'MajorAxis','MinorAxis','Area','Eccentricity','EquivDiameter', ►
'Perimeter','Roundness','Nº Semillas'},'datos','B1');
listamedias=dir('*.TIF_mediamorf.txt');
tablafinalmedias=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
nombres=[listamedias(1,1).name];
for i=1:size(listamedias,1)
variabletemp=load(listamedias(i,1).name);
for j=1:size(variabletemp,1)
nombres=[nombres;listamedias(i,1).name];
end
tablafinalmedias=[tablafinalmedias; variabletemp];
end
tablafinalmedias(1,:)=[];nombres(1,:)=[];
nombres=cellstr(nombres);
xlswrite('MediasMorfología.xls',nombres,'datos','A2');
xlswrite('MediasMorfología.xls',tablafinalmedias,'datos','B2');
xlswrite('MediasMorfología.xls',{'MajorAxis','desv_MajorAxis', ►
'MinorAxis','desv_MinorAxis','Area','desv_Area','Eccentricity', ►
'desv_Eccentricity','EquivDiameter','desv_EquivDiameter', ►
'Perimeter','desv_Perimeter','Roundness','desv_Roundness','Nº ►
Semillas'}, 'datos','B1');
recycle('*morf.txt');disp('Todos los archivos procesados')
end

```

### 3. Software para la clusterización de los colores a partir de una imagen en color CIELAB y su ajuste a elipsoides de color.

Utilizado en el Capítulo 2.

```

tablaCIE=reshape(imagenLAB);
clear imagenLAB
tamanotabla=size(tablaCIE);
for i=1:tamanotabla(1,1);
tablaCIE(i,4)=sqrt((tablaCIE(i,2))^2+(tablaCIE(i,3)^2));
if ((tablaCIE(i,3)>0)&&(tablaCIE(i,2)<0))||((tablaCIE(i,3)<0)&&
(tablaCIE(i,2)<0))tablaCIE(i,5)=(180/pi())*atan(tablaCIE(i,3)/
tablaCIE(i,2))+180;
if (tablaCIE(i,5)<0)
tablaCIE(i,5)=360+tablaCIE(i,5);
end
else
tablaCIE(i,5)=(180/pi())*atan(tablaCIE(i,3)/tablaCIE(i,2));
if (tablaCIE(i,5)<0)
tablaCIE(i,5)=360+tablaCIE(i,5);
end
end
end
imagenCIE=reshape(tablaCIE,tamanoimagen(1),tamanoimagen(2),5);
cocienteCL=imagenCIE(:,:,4)./imagenCIE(:,:,1);
mascaral=(zeros(tamanoimagen(1),tamanoimagen(2)));
for i=1:tamanoimagen(1)
for j=1:tamanoimagen(2)
if cocienteCL(i,j)>2.6
mascaral(i,j)=1;
end
end
end
mascaracruda=imfill(mascaral);
clear mascaral i j
mascara=imerode(mascaracruda,discoerosion);
clear discoerosion cocienteCL mascaracruda
figure(1),imshow(imagen);
figure(2),imshow(logical(mascara));
tablaMASCARA=reshape(mascara,tamanotabla(1),1);
[col,~]=find(tablaMASCARA==1);
pixeles=size(col,1);
tablaCIELAB=zeros(pixeles,5);
for pixel=1:pixeles
for variable=1:5
tablaCIELAB(pixel,variable)=tablaCIE(col(pixel,1),variable);
end
end
reduccion=1000;
submuestreo=unidrnd(pixeles,1,reduccion);
entrada=zeros(reduccion,size(tablaCIELAB,2));
for i=1:reduccion
entrada(i,:)=tablaCIELAB(submuestreo(1,i),:);
end
tablaaab=entrada(:,2:3);
[C]=subclust(tablaaab,0.5,[],[3 0.85 0.15 0]);
clusters=size(C,1);

```

---

```

tablaAB=tablaCIELAB(:,2:3);
[IDX]=kmeans(tablaAB,clusters,'distance','cityblock','start',C,►
'Replicates',1);
[pixelROI(:,1),pixelROI(:,2)]=find(mascara==1);
imagenIDX=zeros(tamanoimagen(1),tamanoimagen(2));
for pixel=1:pixeles
imagenIDX(pixelROI(pixel,1),pixelROI(pixel,2))=IDX(pixel);
end
RGB=label2rgb(imagenIDX);
figure(2),imshow(RGB);
frecuencias=tabulate(reshape(IDX,pixeles,1));
for i=1:clusters
C(i,3)=frecuencias(i,1);
C(i,4)=frecuencias(i,3);
end
C=sortrows(C,-4);
for i=1:clusters
if C(i,4)>=10
nuevaC(i,1:2)=C(i,1:2);
end
end
clusters=size(nuevaC,1);
clear IDX
[IDX]=kmeans(tablaAB,clusters,'distance','cityblock','start',nuevaC,►
'Replicates',1);
imagenIDX=zeros(tamanoimagen(1),tamanoimagen(2));
for pixel=1:pixeles
imagenIDX(pixelROI(pixel,1),pixelROI(pixel,2))=IDX(pixel);
end
RGB=label2rgb(imagenIDX);
figure(2),imtool(RGB);
for numerocluster=1:clusters
nombre=[num2str(numerocluster),'_medias_cluster.txt'];
entradas=find(IDX==numerocluster);
tamanocluster=size(entradas);
cluster(:,1)=tablaCIELAB(entradas(:,1),:);
tablaMCDM=zeros(tamanocluster(1),1);
mediaMCDM=[mean(cluster(:,1)) mean(cluster(:,2)) mean(cluster(:,3))];
for i=1:tamanocluster(1)
tablaMCDM(i,1)=sqrt((cluster(i,1)-mediaMCDM(1))^2+(cluster(i,2)- ►
mediaMCDM(2))^2+(cluster(i,3)-mediaMCDM(3))^2);
end
MCDM=mean(tablaMCDM(:,:));
clear i mediaMCDM tablaMCDM
ab=cluster(:,2:3);
media=[0 0];
media(1,1)=mean(ab(:,1));
media(1,2)=mean(ab(:,2));
[angulo]=robustfit(ab(:,1),ab(:,2));
angulo=atan(angulo(2));
rotacion=zeros(2,2);
rotacion(1,1)=cos(angulo);
rotacion(1,2)=-sin(angulo);
rotacion(2,1)=sin(angulo);
rotacion(2,2)=cos(angulo);
ab_rotada=zeros(tamanocluster(1),2);
for i=1:tamanocluster(1)
ab_rotada(i,:)=ab(i,:)*rotacion;
end

```

---

---

```

desviacion=[0 0];
desviacion(1)=std(ab_rotada(:,1));
desviacion(2)=std(ab_rotada(:,2));
dimensiones=[0 0];
if desviacion(1)>desviacion(2)
dimensiones(1)=desviacion(1)*2;
dimensiones(2)=desviacion(2)*2;
else
dimensiones(1)=desviacion(2)*2;
dimensiones(2)=desviacion(1)*2;
end
if radtodeg(angulo)>180
angulocorregido=radtodeg(angulo)-180;
else
angulocorregido=radtodeg(angulo);
end
datoselipse=[media(1) media(2) dimensiones(1) dimensiones(2) ▶
dimensiones(1)/dimensiones(2) angulocorregido];
media=[mean(cluster(:,1)) std(cluster(:,1)) mean(cluster(:,2)) ▶
std(cluster(:,2)) mean(cluster(:,3)) std(cluster(:,3)) ▶
mean(cluster(:,4)) std(cluster(:,4)) mean(cluster(:,5)) ▶
std(cluster(:,5)) MCDM datoselipse];
save(nombre, 'media', '-ASCII');
clear cluster ab ab_rotada angulos entradas media datoselipse ▶
rotacion desviacion dimensiones tamanocluster MCDM angulo ▶
angulocorregido i
end
listamedias=dir('*_medias_cluster.txt');
tablafinalmedias=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
nombres=[listamedias(1,1).name];
for i=1:size(listamedias,1)
variabletemp=load(listamedias(i,1).name);
for j=1:size(variabletemp,1)
nombres=[nombres;listamedias(i,1).name];
end
tablafinalmedias=[tablafinalmedias; variabletemp];
end
tablafinalmedias(1,:)=[];nombres(1,:)=[];
nombres=cellstr(nombres);
xlswrite('Color.xlsx',nombres,'medias','A2');
xlswrite('Color.xlsx',tablafinalmedias,'medias','B2');
xlswrite('Color.xlsx',{ 'L*', 'desv_L*', 'a*', 'desv_a*', 'b*', ▶
'desv_b*', 'C*ab', 'desv_C*ab', 'hab', 'desv_hab', 'MCDM', ▶
'centroide a', 'centroide b', 'alto', 'ancho', 'cociente', ▶
'angulo' }, 'medias', 'B1');
recycle('*_medias_cluster.txt');
[nubeIDX]=kmeans(tablaab,clusters, 'distance', 'cityblock', 'start', ▶
nuevaC, 'Replicates', 1);
figure(3), scatter(tablaab(:,1), tablaab(:,2), 3, nubeIDX, 'filled');
for i=1:clusters
angulo=deg2rad(tablafinalmedias(i,17));
pdeellip(tablafinalmedias(i,12), tablafinalmedias(i,13), ▶
tablafinalmedias(i,14), tablafinalmedias(i,15), angulo);
clear angulo
end

```

---

---

#### 4. Software para la extracción de los espectros de reflectancia de imágenes hiperespectrales.

Utilizado en el Capítulo 4.

```
function average=HyperspectralSeeds(folder)
cd(folder);
disp(strcat('Folder ',[1],folder,' loaded.'))
filelist=dir('*-1.dat');
sizelist=(size(filelist));
wavelengths=( [▶
930.7100220, 1091.119995, 1251.819946, 1412.829956, 1574.140015, ▶
934.0499880, 1094.459961, 1255.170044, 1416.189941, 1577.510010, ▶
937.3900150, 1097.810059, 1258.520020, 1419.550049, 1580.869995, ▶
940.7299800, 1101.150024, 1261.880005, 1422.900024, 1584.229980, ▶
944.0700070, 1104.5, 1265.229980, 1426.260010, 1587.599976, ▶
947.4099730, 1107.839966, 1268.579956, 1429.619995, 1590.959961, ▶
950.75, 1111.189941, 1271.930054, 1432.979980, 1594.329956, ▶
954.0900270, 1114.530029, 1275.280029, 1436.339966, 1597.689941, ▶
957.4299930, 1117.880005, 1278.640015, 1439.689941, 1601.060059, ▶
960.7700200, 1121.219971, 1281.989990, 1443.050049, 1604.420044, ▶
964.0999760, 1124.569946, 1285.339966, 1446.410034, 1607.790039, ▶
967.4400020, 1127.920044, 1288.689941, 1449.770020, 1611.150024, ▶
970.7899780, 1131.260010, 1292.050049, 1453.130005, 1614.520020, ▶
974.1300050, 1134.609985, 1295.400024, 1456.489990, 1617.880005, ▶
977.4699710, 1137.959961, 1298.750000, 1459.849976, 1621.250000, ▶
980.8099980, 1141.300049, 1302.099976, 1463.209961, 1624.609985, ▶
984.1500240, 1144.650024, 1305.459961, 1466.569946, 1627.979980, ▶
987.4899900, 1148, 1308.810059, 1469.930054, 1631.339966, ▶
990.8300170, 1151.339966, 1312.160034, 1473.290039, 1634.709961, ▶
994.1699830, 1154.689941, 1315.520020, 1476.650024, 1638.079956, ▶
997.5100100, 1158.040039, 1318.869995, 1480.010010, 1641.439941, ▶
1000.849976, 1161.390015, 1322.229980, 1483.369995, 1644.810059, ▶
1004.190002, 1164.729980, 1325.579956, 1486.729980, 1648.180054, ▶
1007.530029, 1168.079956, 1328.930054, 1490.089966, 1651.540039, ▶
1010.880005, 1171.430054, 1332.290039, 1493.449951, 1654.910034, ▶
1014.219971, 1174.780029, 1335.640015, 1496.810059, 1658.280029, ▶
1017.559998, 1178.130005, 1339, 1500.170044, 1661.640015, ▶
1020.900024, 1181.479980, 1342.349976, 1503.530029, 1665.010010, ▶
1024.239990, 1184.819946, 1345.709961, 1506.890015, 1668.380005, ▶
1027.589966, 1188.170044, 1349.060059, 1510.250000, 1671.750000, ▶
1030.930054, 1191.520020, 1352.420044, 1513.609985, 1675.109985, ▶
1034.270020, 1194.869995, 1355.770020, 1516.979980, 1678.479980, ▶
1037.609985, 1198.219971, 1359.130005, 1520.339966, 1681.849976, ▶
1040.959961, 1201.569946, 1362.479980, 1523.699951, 1685.219971, ▶
1044.300049, 1204.920044, 1365.839966, 1527.060059, 1688.589966, ▶
1047.640015, 1208.270020, 1369.189941, 1530.420044, 1691.949951, ▶
1050.989990, 1211.619995, 1372.550049, 1533.780029, 1695.319946, ▶
1054.329956, 1214.969971, 1375.910034, 1537.150024, 1698.689941, ▶
1057.670044, 1218.319946, 1379.260010, 1540.510010, 1702.060059, ▶
1061.020020, 1221.670044, 1382.619995, 1543.869995, 1705.430054, ▶
1064.359985, 1225.020020, 1385.969971, 1547.229980, 1708.800049, ▶
1067.699951, 1228.369995, 1389.329956, 1550.599976, 1712.170044, ▶
1071.050049, 1231.719971, 1392.689941, 1553.959961, 1715.540039]);
for counter=1:sizelist(1,1)
hypercube=multibandread(filelist(counter,1).name,[450 320 ▶
240],'float32',0,'bil','ieee-le');
sizecube=size(hypercube);
```

---

---

```
mask=imsubtract(hypercube(:,:,65),hypercube(:,:,160));
binarymask=zeros(sizecube(1,1),sizecube(1,2));
for i=1:sizecube(1,1)
for j=1:sizecube(1,2)
if mask(i,j)>0.20
binarymask(i,j)=1;
else
binarymask(i,j)=0;
end
end
end
clear mascara i j
hypercubeseg=hypercube;
for k=1:sizecube(1,3)
hypercubeseg(:,:,k)=hypercube(:,:,k).*binarymask(:,:,);
end
clear k
table=reshape(hypercubeseg,sizecube(1,1)*sizecube(1,2),sizecube(1,3));
orderedtable=sortrows(table,-1);
clear table
sizetable=size(orderedtable);
i=1;
while i<sizetable(1,1)
if orderedtable(i,1)==0
orderedtable(i:sizetable(1,1),:)=[];
i=sizetable(1,1);
end
i=i+1;
end
average=zeros(1,sizecube(1,3));
for i=1:sizecube(1,3)
average(1,i)=mean(orderedtable(:,i));
end
tofile=strcat(filelist(counter,1).name,'_averageNOeroded.txt');
save(tofile,'average','-ASCII');
clear hypercubeseg orderedtable mask average sizetable i
diskerode=strel('disk',1);
masktempl=imerode(binarymask,diskerode);
masktemp2=imerode(masktempl,diskerode);
clear masktempl diskerode
hypercubeseg=hypercube;
for k=1:sizecube(1,3)
hypercubeseg(:,:,k)=hypercube(:,:,k).*masktemp2(:,:,);
end
clear k
table=reshape(hypercubeseg,sizecube(1,1)*sizecube(1,2),sizecube(1,3));
orderedtable=sortrows(table,-1);
clear table
sizetable=size(orderedtable);
i=1;
while i<sizetable(1,1)
if orderedtable(i,1)==0
orderedtable(i:sizetable(1,1),:)=[];
i=sizetable(1,1);
end
i=i+1;
end
average=zeros(1,sizecube(1,3));
for i=1:sizecube(1,3)
```

---

---

```

average(1,i)=mean(orderedtable(:,i));
end
tofile=strcat(filelist(counter,1).name, '_averageERODED.txt');
save(tofile, 'average', '-ASCII');
clear hypercube average hypercubeseg i orderedtable sizecube sizetable
clear binarymask masktemp2
disp(strcat('Spectrum of',[1]),filelist(counter,1).name, ' DONE!'))
end
listtxt=dir('*averageNOeroded.txt');
finaltable=zeros(1, size((wavelengths),(2)));
names=[listtxt(1,1).name];
for i=1:size(listtxt,1)
temporalvariable=load(listtxt(i,1).name);
for j=1:size(temporalvariable,1)
names=[names; listtxt(i,1).name];
end
finaltable=[finaltable; temporalvariable];
end
finaltable(1,:)=[];names(1,:)=[];
names=cellstr(names);
xlswrite('Seed_AverageSpectra.xls',names,'Spread_NO_eroded','A2');
xlswrite('Seed_AverageSpectra.xls',finaltable,'Spread_NO_eroded',►
'B2');
xlswrite('Seed_AverageSpectra.xls',{'Wavelength'},'Spread_NO_eroded'►
,'A1');
xlswrite('Seed_AverageSpectra.xls',wavelengths,'Spread_NO_eroded',►
'B1');
listtxt=dir('*averageERODED.txt');
finaltable=zeros(1, size((wavelengths),(2)));
names=[listtxt(1,1).name];
for i=1:size(listtxt,1)
temporalvariable=load(listtxt(i,1).name);
for j=1:size(temporalvariable,1)
names=[names; listtxt(i,1).name];
end
finaltable=[finaltable; temporalvariable];
end
finaltable(1,:)=[];names(1,:)=[];
names=cellstr(names);
xlswrite('Seed_AverageSpectra.xls',names,'Spread_ERODED','A2');
xlswrite('Seed_AverageSpectra.xls',finaltable,'Spread_ERODED','B2');
xlswrite('Seed_AverageSpectra.xls',{'Wavelength'},'Spread_ERODED',►
'A1');
xlswrite('Seed_AverageSpectra.xls',wavelengths,'Spread_ERODED','B1');
clear filelist sizelist;
filelist=dir('*-2.dat');
sizelist=(size(filelist));
for counter=1:sizelist(1,1)
hypercube=multibandread(filelist(counter,1).name,[200 320 240], ►
'float32',0,'bil','ieee-le');
sizecube=size(hypercube);
mask=imsubtract(hypercube(:,:,65),hypercube(:,:,160));
binarymask=zeros(sizecube(1,1),sizecube(1,2));
for i=1:sizecube(1,1)
for j=1:sizecube(1,2)
if mask(i,j)>0.40
binarymask(i,j)=1;
else
binarymask(i,j)=0;

```

---

---

```
end
end
end
clear mascara i j
hypercubeseg=hypercube;
for k=1:sizecube(1,3)
hypercubeseg(:,:,k)=hypercube(:,:,k).*binarymask(:,:,);
end
clear k binarymask
table=reshape(hypercubeseg,sizecube(1,1)*sizecube(1,2),sizecube(1,3));
orderedtable=sortrows(table,-1);
clear table
sizetable=size(orderedtable);
i=1;
while i<sizetable(1,1)
if orderedtable(i,1)==0
orderedtable(i:sizetable(1,1),:)=[];
i=sizetable(1,1);
end
i=i+1;
end
average=zeros(1,sizecube(1,3));
for i=1:sizecube(1,3)
average(1,i)=mean(orderedtable(:,i));
end
tofile=strcat(filelist(counter,1).name,'_HOLDER.txt');
save(tofile,'average','-ASCII');
clear hypercubeseg orderedtable mask average sizetable i
disp(strcat('Spectrum of',[1],filelist(counter,1).name,' DONE!'))
end
listtxt=dir('*HOLDER.txt');
finaltable=zeros(1,size((wavelengths),(2)));
names=[listtxt(1,1).name];
for i=1:size(listtxt,1)
temporalvariable=load(listtxt(i,1).name);
for j=1:size(temporalvariable,1)
names=[names; listtxt(i,1).name];
end
finaltable=[finaltable; temporalvariable];
end
finaltable(1,:)=[];names(1,:)=[];
%% Exportar a Excel
names=cellstr(names);
xlswrite('Seed_AverageSpectra.xls',names,'HOLDER','A2');
xlswrite('Seed_AverageSpectra.xls',finaltable,'HOLDER','B2');
xlswrite('Seed_AverageSpectra.xls',{'Wavelength'},'HOLDER','A1');
xlswrite('Seed_AverageSpectra.xls',wavelengths,'HOLDER','B1');
recycle('*averageNOeroded.txt');
recycle('*averageERODED.txt');
recycle('*HOLDER.txt');
clear
pack
disp(strcat('Process completed!'))
```

---



---

## 5. Software para la extracción de los espectros de reflectancia y absorbancia de imágenes hiperespectrales.

Utilizado en el Capítulo 5.

```
function tablafinal=HiperespectralSemillas▶  
(directorio,bandaini,bandafin,columnas,filas);  
tic;  
WL=[884.380005, 887.630005, 890.880005, 894.130005, 897.369995, ▶  
900.619995, 903.869995, 907.119995, 910.369995, 913.619995, ▶  
916.869995, 920.119995, 923.369995, 926.619995, 929.869995, ▶  
933.119995, 936.369995, 939.619995, 942.869995, 946.119995, ▶  
949.369995, 952.619995, 955.869995, 959.119995, 962.369995, ▶  
965.630005, 968.880005, 972.130005, 975.380005, 978.630005, ▶  
981.890015, 985.140015, 988.390015, 991.640015, 994.900024, ▶  
998.150024, 1001.400024, 1004.659973, 1007.909973, 1011.159973, ▶  
1014.419983, 1017.669983, 1020.919983, 1024.180054, 1027.430054, ▶  
1030.689941, 1033.939941, 1037.199951, 1040.449951, 1043.709961, ▶  
1046.959961, 1050.219971, 1053.469971, 1056.729980, 1059.979980, ▶  
1063.239990, 1066.500000, 1069.750000, 1073.010010, 1076.270020, ▶  
1079.520020, 1082.780029, 1086.040039, 1089.290039, 1092.550049, ▶  
1095.810059, 1099.069946, 1102.319946, 1105.579956, 1108.839966, ▶  
1112.099976, 1115.359985, 1118.609985, 1121.869995, 1125.130005, ▶  
1128.390015, 1131.650024, 1134.910034, 1138.170044, 1141.430054, ▶  
1144.689941, 1147.949951, 1151.209961, 1154.469971, 1157.729980, ▶  
1160.989990, 1164.250000, 1167.510010, 1170.770020, 1174.030029, ▶  
1177.290039, 1180.550049, 1183.810059, 1187.069946, 1190.339966, ▶  
1193.599976, 1196.859985, 1200.119995, 1203.380005, 1206.650024, ▶  
1209.910034, 1213.170044, 1216.430054, 1219.699951, 1222.959961, ▶  
1226.219971, 1229.489990, 1232.750000, 1236.010010, 1239.280029, ▶  
1242.540039, 1245.810059, 1249.069946, 1252.329956, 1255.599976, ▶  
1258.859985, 1262.130005, 1265.390015, 1268.660034, 1271.920044, ▶  
1275.189941, 1278.459961, 1281.719971, 1284.989990, 1288.250000, ▶  
1291.520020, 1294.790039, 1298.050049, 1301.319946, 1304.579956, ▶  
1307.849976, 1311.119995, 1314.390015, 1317.650024, 1320.920044, ▶  
1324.189941, 1327.459961, 1330.719971, 1333.989990, 1337.260010, ▶  
1340.530029, 1343.800049, 1347.069946, 1350.329956, 1353.599976, ▶  
1356.869995, 1360.140015, 1363.410034, 1366.680054, 1369.949951, ▶  
1373.219971, 1376.489990, 1379.760010, 1383.030029, 1386.300049, ▶  
1389.569946, 1392.839966, 1396.109985, 1399.380005, 1402.650024, ▶  
1405.930054, 1409.199951, 1412.469971, 1415.739990, 1419.010010, ▶  
1422.280029, 1425.560059, 1428.829956, 1432.099976, 1435.369995, ▶  
1438.650024, 1441.920044, 1445.189941, 1448.469971, 1451.739990, ▶  
1455.010010, 1458.290039, 1461.560059, 1464.829956, 1468.109985, ▶  
1471.380005, 1474.660034, 1477.930054, 1481.199951, 1484.479980, ▶  
1487.750000, 1491.030029, 1494.300049, 1497.579956, 1500.859985, ▶  
1504.130005, 1507.410034, 1510.680054, 1513.959961, 1517.239990, ▶  
1520.510010, 1523.790039, 1527.060059, 1530.339966, 1533.619995, ▶  
1536.900024, 1540.170044, 1543.449951, 1546.729980, 1550.010010, ▶  
1553.280029, 1556.560059, 1559.839966, 1563.119995, 1566.400024, ▶  
1569.670044, 1572.949951, 1576.229980, 1579.510010, 1582.790039, ▶  
1586.069946, 1589.349976, 1592.630005, 1595.910034, 1599.189941, ▶  
1602.469971, 1605.750000, 1609.030029, 1612.310059, 1615.589966, ▶  
1618.869995, 1622.150024, 1625.430054, 1628.709961, 1631.989990, ▶  
1635.280029, 1638.560059, 1641.839966, 1645.119995, 1648.400024, ▶  
1651.680054, 1654.969971, 1658.250000, 1661.530029, 1664.819946, ▶  
1668.099976, 1671.380005, 1674.660034, 1677.949951, 1681.229980, ▶  
1684.510010, 1687.800049, 1691.079956, 1694.369995, 1697.650024, ▶
```

---

```
1700.939941, 1704.219971, 1707.500000, 1710.790039, 1714.069946, ▶
1717.359985];
funciondis=[-211.9780000000000,1430.390000000000;-513.854000000000,▶
-1165.100000000000;538.863000000000,304.260000000000; ▶
-88.0550000000000,-258.900000000000;];
bandasdis=[103;157;82;];
cd(directorio);
listaarchivos(:,1)=dir('*.hdr');
listaarchivos(:,2)=dir('*.dat');
tamanolista=size(listaarchivos);
for contador=1:tamanolista(1);
importheader=importdata(listaarchivos(contador,1).name);
header=importheader.textdata; clear importheader;
A=char(header(4,1));
B=char(header(5,1));
C=char(header(6,1));
samples=str2double(A(10:13));
lines=str2double(B(10:13));
bands=str2double(C(10:13));
clear A B C header
image=multibandread(listaarchivos(contador,2).name,▶
[lines samples bands],'float32',0,'bil','ieee-le');
matrizdis=zeros([lines],[samples],2);
for i=1:lines
for j=1:samples
for condicion=1:2
matrizdis(i,j,condicion)=(image(i,j,103)*funciondis(1,condicion))+ ▶
(image(i,j,157)*funciondis(2,condicion))+(image(i,j,29)* ▶
funciondis(3,condicion))+funciondis(4,condicion);
end
end
end
mascaradis=zeros([lines],[samples]);
for i=1:lines
for j=1:samples
if matrizdis(i,j,2)>matrizdis(i,j,1)
mascaradis(i,j)=1;
end
end
end
[row,col]=find(mascaradis==1);
limitesF=zeros(filas+1,1);
limitesC=zeros(1,columnas+1);
for i=1:filas
limitesF(i+1,1)=quantile(row,i/filas);
end
for j=1:columnas
limitesC(1,j+1)=quantile(col,j/columnas);
end
plantilla=zeros([lines],[samples],2);
for f=1:filas
for i=1:limitesF(filas+1,1)
if i>limitesF(f,1)
plantilla(i,:,1)=f;
end
end
end
for c=1:columnas
for j=1:limitesC(1,columnas+1)
```

---

---

```

if j>limitesC(1,c)
plantilla(:,j,2)=c;
end
end
end
contadormuestra=1;
for h=1:filas
for v=1:columnas
mascarasub=(zeros(lines,samples));
for i=1:lines
for j=1:samples
if plantilla(i,j,1)==h && plantilla(i,j,2)==v && mascaradis(i,j)==1
mascarasub(i,j)=1;
end
end
end
etiqueta=num2str(contadormuestra);
contadormuestra=contadormuestra+1;
[rowSUB,colSUB]=find(mascarasub==1);
puntos=size(colSUB);puntos(2)=[];
espectros=zeros(puntos,256);
for lambda=1:256
for punto=1:puntos
espectros(punto,lambda)=image(rowSUB(punto,1),colSUB(punto,1),lambda);
end
reflectancia(1,lambda)=mean(espectros(:,lambda));
absorbancia(1,lambda)=log10(1/(mean(espectros(:,lambda))));
end
reflec=reflectancia(1,bandaini:1:bandafin);
absorb=absorbancia(1,bandaini:1:bandafin);
guardatrans=strcat(listaarchivos(contador,2).name,'_',etiqueta,►
'_reflectancia.txt');
save(guardatrans,'reflec','-ASCII');
guardaabs=strcat(listaarchivos(contador,2).name,'_',etiqueta,►
'_absorbancia.txt');
save(guardaabs,'absorb','-ASCII');
guardamask=uint8(zeros([lines],[samples],3));
for i=1:lines
for j=1:samples
if mascarasub(i,j)==1;
guardamask(i,j,1)=255;
guardamask(i,j,2)=255;
guardamask(i,j,3)=255;
end
end
end
nombreimagen=strcat(listaarchivos(contador,2).name,'_',etiqueta,►
'_mascara.jpg');
imwrite(guardamask,nombreimagen);
end
end
end
listtxt=dir('*reflectancia.txt');
wavelengths=zeros(1,bandafin-bandaini+1);
for i=1:bandafin-bandaini+1
wavelengths(1,i)=WL(1,i+bandaini-1);
end
tablafinal=zeros(1,size((wavelengths),(2)));
names=[listtxt(1,1).name];

```

---

---

```
for i=1:size(listtxt,1)
temporalvariable=load(listtxt(i,1).name);
for j=1:size(temporalvariable,1)
names=[names; listtxt(i,1).name];
end
tablafinal=[tablafinal; temporalvariable];
end
tablafinal(1,:)=[];names(1,:)=[];
names=cellstr(names);
xlswrite('HiperespectrosSemillas.xlsx',names,'Ref','A2');
xlswrite('HiperespectrosSemillas.xlsx',tablafinal,'Ref','B2');
xlswrite('HiperespectrosSemillas.xlsx',{'Wavelength'},'Ref','A1');
xlswrite('HiperespectrosSemillas.xlsx',wavelengths,'Ref','B1');
recycle('*reflectancia.txt');
listtxt=dir('*absorbancia.txt');
wavelengths=zeros(1,bandafin-bandaini+1);
for i=1:bandafin-bandaini+1
wavelengths(1,i)=WL(1,i+bandaini-1);
end
tablafinal=zeros(1,size((wavelengths),(2)));
names=[listtxt(1,1).name];
for i=1:size(listtxt,1)
temporalvariable=load(listtxt(i,1).name);
for j=1:size(temporalvariable,1)
names=[names; listtxt(i,1).name];
end
tablafinal=[tablafinal; temporalvariable];
end
tablafinal(1,:)=[];names(1,:)=[];
names=cellstr(names);
xlswrite('HiperespectrosSemillas.xlsx',names,'Abs','A2');
xlswrite('HiperespectrosSemillas.xlsx',tablafinal,'Abs','B2');
xlswrite('HiperespectrosSemillas.xlsx',{'Wavelength'},'Abs','A1');
xlswrite('HiperespectrosSemillas.xlsx',wavelengths,'Abs','B1');
recycle('*absorbancia.txt');
tiempo_final=toc
end
```

---



