



Potential of VIS-NIR hyperspectral imaging and chemometric methods to identify similar cultivars of nectarine



Sandra Munera^a, Jose Manuel Amigo^{b,c}, Nuria Aleixos^d, Pau Talens^e, Sergio Cubero^a, José Blasco^{a,*}

^a Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), Carretera CV-315, Km 10.7, 46113 Moncada, Spain

^b Department of Food Sciences, Faculty of Science, University of Copenhagen, Rolighedsvej 30, Frederiksborg C DK-1958, Denmark

^c Department of Fundamental Chemistry, Federal University of Pernambuco, Av. Prof. Moraes Rego, 1235, Cidade Universitária, Recife, Brazil

^d Departamento de Ingeniería Gráfica, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain

^e Departamento de Tecnología de Alimentos, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain

ARTICLE INFO

Article history:

Received 31 July 2017

Received in revised form

29 October 2017

Accepted 30 October 2017

Available online 1 November 2017

Keywords:

Stone fruit

Quality control

Cultivar discrimination

Non-destructive

PLS-DA

Colour analysis

Hyperspectral image

ABSTRACT

Product inspection is essential to ensure good quality and to avoid fraud. New nectarine cultivars with similar external appearance but different physicochemical properties may be mixed in the market, causing confusion and rejection among consumers, and consequently affecting sales and prices. Hyperspectral reflectance imaging in the range of 450–1040 nm was studied as a non-destructive method to differentiate two cultivars of nectarines with a very similar appearance but different taste. Partial least squares discriminant analysis (PLS-DA) was used to develop a prediction model to distinguish intact fruits of the cultivars using pixel-wise and mean spectrum approaches, and then the model was projected onto the complete surface of fruits allowing visual inspection. The results indicated that mean spectrum of the fruit was the most accurate method, a correct discrimination rate of 94% being achieved. Wavelength selection reduced the dimensionality of the hyperspectral images using the regression coefficients of the PLS-DA model. An accuracy of 96% was obtained by using 14 optimal wavelengths, whereas colour imaging and a trained inspection panel achieved a rate of correct classification of only 57% of the fruits.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The surface area of the land devoted to the planting of peaches and nectarines (*Prunus persica* L. Batsch) in the EU was around 232 000 ha in 2015/16, with a production of nearly 3.7 million tons of fruit. Spain is the main producer with around 1.4 tons, which accounts for almost 40% of the total EU peach and nectarine production (USDA, 2016). Due to the importance of nectarine (*Prunus persica* L. Batsch var. *nucipersica*) production, it is one of the fruits to which most effort has been devoted by plant breeders in recent years in order to improve agronomic performance, and enhanced fruit appearance and quality (Reig, Alegre, Gatiús, & Iglesias, 2013). This fact has resulted in a significant increase in the number of new cultivars available to fruit growers. These cultivars are similar in appearance but present different sensory properties and therefore

different acceptance by the consumer (Iglesias & Echeverría, 2009). In this context, one of the most widely accepted and cultivated nectarine cultivars in Europe is 'Big Top' due to its presentation, size, sweet taste and low acidity (Echeverría, Cantín, Ortiz, López, & Graell, 2015). However, a stagnation of nectarine consumption is occurring owing to early harvesting, which leads to flavourless fruits being offered with excessive F or irregular quality (Iglesias & Echeverría, 2009). These authors also point out the lack of an adequate identification of the product in the market. The mixture of sweet and acid cultivars on the shelf could lead to consumer rejection, which in turn might affect sales and prices.

The internal quality assessment of stone fruits has traditionally been performed by destructive methods, which are contaminating, time-consuming and only a few samples per batch can be monitored (Pérez-Marín, Sánchez, Paz, González-Dugo, & Soriano, 2011). Moreover, there is an important lack of classification tools for differentiating cultivars that are very similar to one another. There is therefore a strong need to develop non-destructive and instantaneous methodologies that allow the correct identification of the

* Corresponding author.

E-mail address: blasco_josiva@gva.es (J. Blasco).

Abbreviations

ANOVA	analysis of variance
CCD	charge-coupled device
CV	cross validation
EU	European Union
F	firmness
LV	latent variables
NIR	near infrared
PC	principal component
PCA	principal component analysis
PLS-DA	partial least square discriminant analysis
RGB	red, green, blue
SC	skin colour
SNV	standard normal variate
TA	titratable acidity
TSS	total soluble solids
VIS	visible

cultivar in the postharvest stage.

Hyperspectral imaging is a computer vision technique which combines conventional two-dimensional digital imagery with spectroscopy to detect spectral features in regions of the electromagnetic spectrum such as the ultraviolet, NIR or infrared regions (Lorente et al., 2012). This technique is starting to be used as a scientific tool for quality assurance of a wide range of food including bakery products (Erkinbaev, Henderson, & Paliwal, 2017; Verdú et al., 2016), meat (Feng, Makino, Oshita, & García-Martín, 2018; Iqbal, Sun, & Allen, 2014), or vegetables (López-Maestresalas et al., 2016). Fruits are of major interest for the use of this technology in the food industry (Keresztes, Goodarzi, & Saeys, 2016; Munera, Amigo et al., 2017). However, due to the high importance of other fruits such as citrus or apples, few scientific studies have been done for quality control of stone fruit quality assessment using hyperspectral imaging. Herrero-Langreo, Lunadei, Lleó, Diezma, and Ruiz-Altisent (2011) assessed the ripeness of peaches by using multispectral indexes. Lu and Peng (2006) assessed the F of peaches and Zhu, Lin, Nie, Wu, and Chen (2016) obtained F distribution maps inside the peach pulp, while Zhang et al. (2015), Li et al. (2016), Pan et al. (2016) and Sun et al. (2017) detected different types of defects and injuries, including decay. Regarding nectarine, Huang et al. (2015) used the same technique to detect defective features and Munera, Besada et al. (2017) to monitor its ripeness.

Hyperspectral imaging generates a huge amount of redundant and frequently highly correlated data that need to be processed (Vélez-Rivera et al., 2014). To handle such an amount of data and extract the useful information, it must be assisted by chemometric methods. These methods connect chemical measurements with the essential spectral information in order to classify and/or quantify important characteristics. PCA is one of the most popular methods commonly used both to reduce the dimensionality of data and to obtain an overview of all the relevant information in the dataset. It is an unsupervised projection method which summarises data by forming new independent linear combinations of the original variables (Jolliffe, 2002).

PLS-DA is a variant of PLS regression in which the independent variable is categorical, expressing the class membership of the samples. It is performed in order to sharpen the separation between groups of observations by maximising the covariance between the spectra and the independent variable such that a

maximum separation among classes is obtained. Furthermore, it is commonly used to understand which variables contain the discriminating information (Lorente et al., 2012). Some examples of the use of this method include the detection of decay lesions in citrus fruits (Folch-Fortuny, Prats-Montalbán, Cubero, Blasco, & Ferrer, 2016), classification of oat kernels (Serranti, Cesare, Marini, & Bonifazi, 2013), the classification of edible fennel heads based on the harvest time (Amodio, Capotorto, Chaudhry, & Colelli, 2017), and the examination of aflatoxin on corn kernels (Kandpal, Lee, Kim, Bae, & Cho, 2015).

In this paper, we put forward a novel approach based on VIS-NIR hyperspectral imaging and chemometric methods to develop statistical predictive models capable of distinguishing cultivars of nectarines with a very similar appearance but different taste. Previous studies have been conducted to differentiate among nectarine cultivars using colour images (Font et al., 2014). However, they use fruits with clearly different appearance. In this work, 'Diamond Ray' and 'Big Top' cultivars have been used due to their similar skin and flesh appearance. Furthermore, these cultivars are grown and marketed at the same time and become a problem for producers when they are mixed, either accidentally or intentionally, in the market.

In addition, using the spectral and spatial information provided by the hyperspectral images, two approaches are further investigated: the first based on the analysis of the individual spectrum of each pixel and the second based on the mean spectrum of each fruit. Finally, visualisation of the result of the classification model over the images of nectarines is proposed to establish a practical tool for nectarine classification in the packing houses.

2. Material and methods

2.1. Fruit samples

Nectarines cv. 'Diamond Ray' and 'Big Top' were selected as reference cultivars of sweet and acid cultivars, respectively (Reig, Iglesias, & Echeverría, 2009), due to their similar skin and flesh appearance. These two cultivars are difficult to distinguish by the naked eye, which is problematic for producers when they are mixed in the market.

Fruits were harvested in a commercial orchard in Lerida (Spain) at the commercial maturity stage in the summer season of 2016. A total of 125 fruits of each cultivar without defects or bruises were selected and stored under controlled conditions (1 °C; 90% relative humidity) in order to avoid the further ripening of either cultivar during the experiment.

2.2. Hyperspectral image acquisition and processing

The hyperspectral imaging system consisted of an industrial camera (CoolSNAP ES, Photometrics, AZ, USA), coupled to two liquid-crystal tuneable filters (Varispec VIS-07 and NIR-07, Cambridge Research & Instrumentation, Inc., MA, USA). The camera was configured to acquire images with a size of 1392 × 1040 pixels and a spatial resolution of 0.14 mm/pixel at 60 different wavelengths every 10 nm, in the working spectral range of 450 nm–1040 nm. In order to avoid problems of unfocused images due to the refraction of light across this wide spectral range, the focus was adjusted on the central band of the acquisition interval (740 nm) and the images were captured using lenses capable of covering the whole spectral range without going out of focus (Xenoplan 1.4/23, Schneider Optics, Hauppauge, NY, USA). To optimise the dynamic range of the camera, prevent saturated images and correct the spectral sensitivity of the different elements of the system, a calibration of the integration time of each band was performed by capturing the

averaged grey level of a white reference target (Spectralon 99%, Labsphere, Inc, NH, USA) corresponding to 90% of the dynamic range of the camera.

The scene was illuminated by indirect light from twelve halogen spotlights (37 W) (Eurostar IR Halogen MR16, Ushio America, Inc., CA, USA) powered by direct current (12 V) and arranged equidistant from each other inside a hemispherical aluminium diffuser. The inner surface of the aluminium diffuser was painted white with a rough texture to maximise its reflectivity, the rough texture being applied in order to minimise directional reflections, which could cause bright spots, thus resulting in highly homogeneous light.

The fruits were introduced manually into a fruit holder, with the stem-calyx axis lying horizontal. Two images of each fruit were acquired using customised software developed at IVIA. A total of 250 images of each cultivar were imported into MATLAB R2015a (The MathWorks, Inc. MA, USA) to be pre-processed using the customised toolbox HYPER-Tools (Amigo, Babamoradia, & Elcoroaristizabal, 2015).

The image processing started with the correction of the relative reflectance by using equation (1) (Gat, 2000):

$$\rho_{xy}(x, y, \lambda) = \frac{R_{\text{abs}}}{R_{\text{white}}} = \rho^{\text{Ref}}(\lambda) \frac{R(x, y, \lambda) - R_{\text{black}}(x, y, \lambda)}{R_{\text{white}}(x, y, \lambda) - R_{\text{black}}(x, y, \lambda)} \quad (1)$$

where $\rho^{\text{Ref}}(\lambda)$ is the standard reflectance of the white reference target (99% in this work), $R(x, y, \lambda)$ is the reflectance of the fruit captured by the CCD sensor of the camera, $R_{\text{white}}(x, y, \lambda)$ is the reflectance captured by the CCD of the white reference target, and $R_{\text{black}}(x, y, \lambda)$ is the reflectance captured by the CCD while avoiding any light source in order to quantify the electronic noise of the CCD. The images were then clipped and spatially compressed to reduce the computation time, and a proper removal of the background was performed using K-means clustering. Thus, the relative reflectance spectrum of all the pixels in each fruit image was extracted.

2.3. Colour image acquisition and processing

Before image acquisition, the SC was analysed to obtain the L^* , a^* and b^* colour coordinates (CIELAB colour space) of each fruit, also using a colorimeter (MINOLTACM-700d, Minolta Co. Tokyo, Japan) configured with the standard illuminant D65 and the observer 10°. The SC was obtained as the average of the values of two measurements, one in the blush zone (reddish colour) and another in the ground zone (yellowish colour).

The colour imaging system consisted of a digital camera (EOS 550D, Canon Inc, Japan) arranged inside a square inspection chamber that included a calibrated and uniform illumination system composed of four lamps, each containing two fluorescent tubes BIOLUX 18W/965 (Osram GmbH, Germany) with a colour temperature of 6500 K. The angle between the axis of the lens and the sources of illumination was approximately 45°, and polarising filters were placed in front of the lamps and in the camera lenses to eliminate specular bright spots that could alter the true colour.

The fruits were introduced manually upon a fruit holder, with the stem-calyx axis lying horizontal. Two images were acquired for each fruit, corresponding to each of the two sides delimited by the suture of the fruit. Then, a total of 250 images of each cultivar were imported into customised software developed at IVIA (FoodImage-Inspector v4.0, freely available at <http://www.cofilab.com>, Spain) to analyse the SC and to obtain the percentage of the reddish and yellowish zones on the fruit. This segmentation was based on the Bayes theorem to assign all the pixels in the image to the two classes used in a previous training. The RGB colour coordinates of

the acquired images were converted to the $L^*a^*b^*$ coordinates and then corrected using a colour reference target (ColorChecker Digital SG, X-Rite, MI, USA).

2.4. Visual analysis with trained panel

The panel was composed of five panellists, ages 29–50 years (three male and two female), with expertise in fruit quality and marketing. The panellists were trained using 20 colour images of nectarines of the calibration set (10 from each cultivar chosen at random). A total of 40 colour images of fruits of the validation set (20 from each cultivar chosen at random) were presented with randomised order to each panellist to be classified as belonging to the 'Diamond Ray' or 'Big Top' cultivar.

2.5. Reference analysis

The characterisation of the physicochemical properties of the samples using reference methods was performed immediately after the acquisition of the images. F was registered on opposite sides of the fruits using an XT2 Stable texturometer (MicroSystems Haslemere, UK) equipped with a 6 mm flat plunger. The crosshead speed during the puncture test was 1 mm/s. The maximum force was expressed in Newton (N). Immediately after SC and F measurements, a juice sample was taken from each fruit for TSS and TA measurements. TSS were determined using a digital refractometer RFM330 + VWR (Internacional Eurolab S.L., Barcelona, Spain) at 20 °C and results were expressed as percentage of TSS. TA was determined using a Crison pH-Burette 24 automatic titrator (Crison, Barcelona, Spain) and NaOH 0.5 N, according to standard UNE34211:1981 (AENOR, 1981). The results were expressed as the percentage of malic acid.

The ANOVA was conducted using the software Statgraphics (Manugistics Corp., Rockville, USA) in order to determine significant differences in the physicochemical properties (F, TSS, TA and L^* , a^* and b^* colour coordinates) between cultivars.

2.6. Chemometric methods

To identify both nectarine cultivars with high precision, two approaches were studied for setting up the classification models: i) including in the model the individual spectrum of each pixel in the nectarine image, and ii) using only the mean spectrum of all the pixels corresponding to each fruit. Thus, 512 828 pixel spectra were used in the first approach, and the mean spectra of 500 fruits were used for the second. The data of all the fruits of both cultivars were collected and randomly partitioned into two sets: two thirds of the samples were used to calibrate the models (calibration set) and for cross-validation, while the remaining third was used for independent test prediction (validation set).

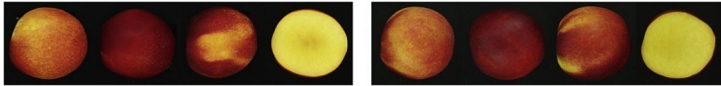
Both the directly acquired spectrum of each pixel and that obtained as an average for each fruit were pre-processed using SNV in order to reduce the physical variability between samples due to light scatter (Rinnan, van den Berg, & Engelsen, 2009). This correction was performed using equation (2):

$$x_{\text{corr}} = \frac{x_{\text{org}} - a_0}{a_1} \quad (2)$$

where x_{corr} and x_{org} are the corrected and raw spectra, respectively, a_0 is the average value of the sample spectrum to be corrected and a_1 is the standard deviation of the sample spectrum.

Later, mean centring was applied to normalise the full spectrum. Multivariate analyses were then performed using the PLS_Toolbox (Eigenvector Research Inc., USA) and the HYPER-Tools toolbox

Table 1
Results of analysis of physicochemical properties of both cultivars of nectarine.

Property		'Diamond Ray'		'Big Top'	
		Mean	SD	Mean	SD
Firmness (N)		33.8 ^a	9.5	34.8 ^a	7.1
Total soluble solids (%)		11.9 ^b	1.6	12.7 ^a	2.3
Titratable acidity (%)		0.7 ^a	0.1	0.4 ^b	0.1
Skin colour by colorimeter	<i>L</i> *	36.9 ^a	6.6	36.5 ^a	6.0
	<i>a</i> *	27.0 ^a	4.2	26.2 ^a	3.9
	<i>b</i> *	13.3 ^a	5.1	13.4 ^a	4.9
Skin colour by imaging	<i>L</i> *	28.0 ^a	8.6	27.0 ^a	8.2
	<i>a</i> *	44.9 ^a	5.4	41.0 ^b	5.5
	<i>b</i> *	27.2 ^a	8.8	24.8 ^b	8.6
	Blush zone (%)	67.0 ^a	21.4	66.3 ^a	18.4
	Ground zone (%)	33.0 ^a	21.4	33.7 ^a	18.4
External and internal appearance					

Different superscript letters in the same row indicate significant differences between cultivars (p-value<0.05). SD = standard deviation.

(Amigo et al., 2015) both working under MATLAB R2015a.

PCA was used to explore the differences between the two cultivars using the pixel and mean spectra of the calibration set previously pre-processed by means of SNV and mean centring. Later, PLS-DA models were built to sort the fruits into one of the two studied cultivars. The models were also calibrated using the pre-processed pixel and mean spectra of the calibration set and tested using only samples of the validation or prediction set.

In order to compare the performance of the hyperspectral imaging with the colour imaging system, a PLS model was also built using the mean value of the *L***a***b** colour coordinates.

A single 10-fold venetian blind CV was used to choose the optimal number of LV as well as to obtain an estimation of the error rate of the models. All models were statistically validated by using the sensitivity, specificity, class error and accuracy (Eqs. (3)–(6)):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$\text{Class error} = 1 - \left(\frac{\text{Sensitivity} + \text{Specificity}}{2} \right) \quad (5)$$

$$\text{Accuracy (\%)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (6)$$

where TP and TN stand for true positive and true negative, respectively, accounting for the samples that have been correctly assigned as belonging (TP), or not belonging (TN), to a specific class. FP and FN stand for false positive and false negative, respectively, accounting for the samples that have been wrongly assigned as belonging (FP), or not belonging (FN), to a specific class.

The ANOVA, using the software Statgraphics, was also conducted in order to determine significant differences in the accuracy of the models.

3. Results and discussion

3.1. Cultivar characterisation

3.1.1. Physicochemical properties

Table 1 shows the results obtained from the reference analysis of the physicochemical properties. F is one of the physicochemical properties commonly used to assess ripeness. In this work, the measures of F obtained for both cultivars showed no statistical differences, which means that they were in a similar stage of ripeness. According to the mean value of F measured for each cultivar, these fruits were considered as being within the group that Valero, Crisosto, and Slaughter (2007) described as 'ready to buy'.

As noted above, the principal difference between these two cultivars is the flavour; i.e. the typical TSS values for 'Big Top' being higher than in 'Diamond Ray' and vice versa for TA. The measured values (Table 1) agreed with Crisosto et al. (2006), who found that 'Diamond Ray' had 0.8% TA and 10.3% TSS. The difference in TSS content between these cultivars may be attributable to the stage of maturity, the season or the production area (Crisosto, 1994). Regarding the 'Big Top' cultivar, Giné-Bordonaba et al. (2014) reported results similar to those in the present study, i.e. 0.3% TA and TSS between 12.2% and 13.5%.

The mean *L**, *a**, and *b** colour coordinates of the SC using the colorimeter were not statistically different between cultivars (Table 1). However, colorimeters measure small regions only, which can be a major limitation in applications where distinguishing the colours all over the sample is of interest. This means that they are not well suited to measuring objects with a heterogeneous colour (Gardner, 2007), such as nectarines of these cultivars. However, a colour camera provides images in which the colours of the pixels are determined individually (Cubero, Aleixos, Moltó, Gómez-Sanchis, & Blasco, 2011), along with their spatial distribution. The analysis of the colour of the nectarines using imaging enable the evaluation of the SC of the different colour zones separately and calculation of the relative distribution (percentage) of reddish or yellowish colour in the whole fruit.

Using this percentage, a mean value of the *L**, *a**, *b** coordinates was calculated from the images. On average, a reddish colour was present on 67% of the fruit surface and a yellowish colour on 33% in both cultivars (Table 1). Even so, the mean colour using imaging indicated that the *a** and *b** scores were statistically different in the two cultivars, i.e. both were higher in 'Diamond Ray'. However, the

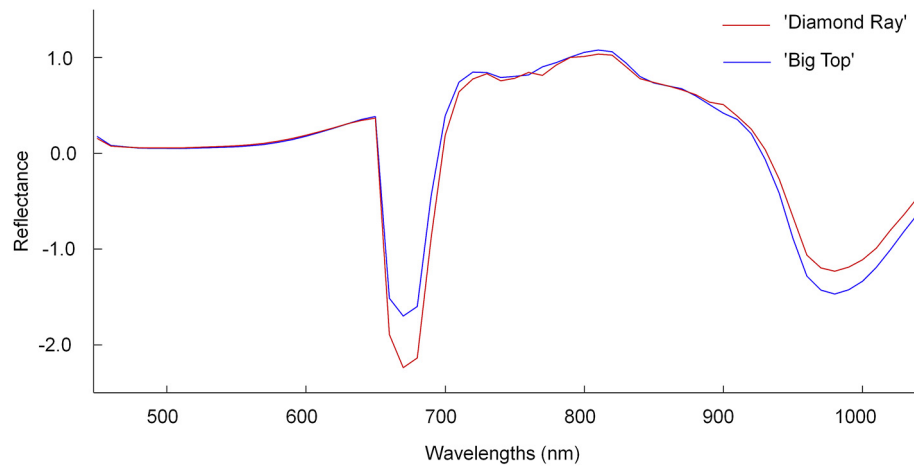


Fig. 1. Mean hyperspectral image spectra of 'Diamond Ray' and 'Big Top' cultivars.

differences were too small to be detected visually by the human eye, especially during a rapid fruit-sorting process.

3.1.2. Spectral analysis

Differences between cultivars were observed in their hyperspectral spectra (Fig. 1). The pre-processed (SNV) mean spectra of the two cultivars followed a similar spectral pattern but had clear differences at specific wavelengths.

In the VIS region, no apparent differences could be visualised in the range between 400 and 600 nm where carotenoids are present. In contrast, the 'Big Top' cultivar had lower reflectance (higher absorbance) than 'Diamond Ray' near 680 nm, which is associated with chlorophylls (Rajkumar, Wang, Elmasry, Raghavan, & Gariepy, 2012), suggesting a higher content of this molecule. This agrees with the differences in the values of a^* and b^* found in the colour analysis (Table 1).

In the NIR region, the absorption bands for acids and sugars are usually found around 800 nm and 840 nm respectively, attributable to the hydroxyl groups of these compounds (Malegori, Nascimento, Tonetto de Freitas, Pimentel, & Casiraghi, 2017; Yang, Sun, Pu, Wang, & Zhu, 2015). However, only small differences are usually observable due to the water absorption bands which dominate the spectrum (Nicolai et al., 2007). In this region, the main differences observed in the spectra were at wavelengths above 850 nm and, in

particular, around 970 nm, where Lu and Peng (2006) described a peak associated with water absorption, which in this case was more pronounced in 'Big Top' nectarines (Fig. 1).

3.2. Overview of the spectral data

A PCA was performed in order to obtain an overview of the distribution of the spectral data information from the samples of both cultivars. The PCA results from the individual pixel spectra and the mean spectra of each fruit are shown in Figs. 2 and 3, respectively.

Forty samples of each cultivar were randomly selected to provide individual pixel spectra and this data was used to generate a score image plot. The first two PCs explained 87.8% of the total variance (76.5% and 11.3%, respectively). The variations in the colour within each fruit showed the distribution or content of the biochemical constituents. A possible trend was discerned in PC2, where pixels with low values (dark blue) were found mostly in 'Big Top' samples; however, there was little difference in individual fruit spectra of the 'Diamond Ray' and 'Big Top' cultivars.

In the PCA of the mean spectra of the calibration set, the first two PCs (Fig. 3A) explained 93.3% of the variance (81.4% and 11.9%, respectively). The ellipses for the two cultivars appeared distinct, but discrimination between them was not possible because of

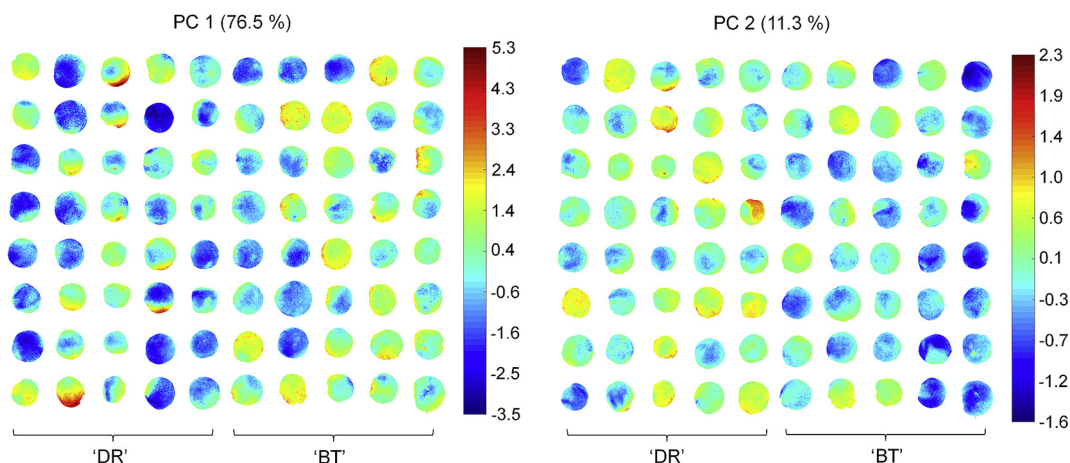


Fig. 2. Score image of the two first PC of the PCA model using pixel spectra of 40 fruits of each cultivar from the calibration set.

Key for Fig. 2: The percentages indicate the explained variance (87.8% of the total variance). The variations in the colour in both score plots show features linked to the distribution or content of the biochemical constituents in each fruit and cultivar. 'DR' = 'Diamond Ray'; 'BT' = 'Big Top'.

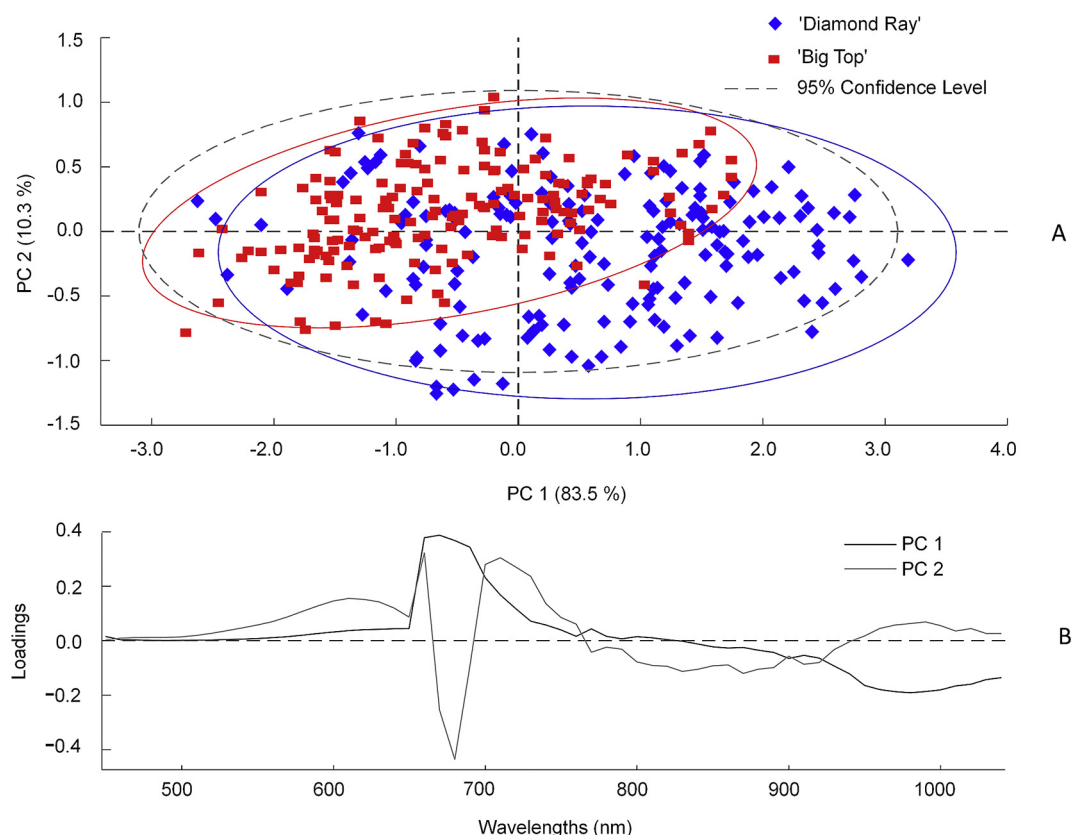


Fig. 3. Score (A) and loadings plot (B) of the PCA of the mean spectra of the calibration set.

Table 2

Cultivar discrimination using the pixel spectrum approach.

#V	#LV	Set	Class	Sensitivity	Specificity	Error	Accuracy (%)
60	5	Calibration	'DR'	0.83	0.86	0.15	84.8
			'BT'	0.86	0.83		
		Cross Validation	'DR'	0.83	0.86	0.15	84.8
			'BT'	0.86	0.83		
		Validation pixel	'DR'	0.79	0.89	0.16	83.8
			'BT'	0.89	0.79		
		Validation object	'DR'	0.78	0.91	0.16	84.4
			'BT'	0.91	0.78		

V=Variables; LV = Latent variables; 'DR' = Diamond Ray; 'BT' = 'Big Top'.

overlap (Fig. 3A).

Although the loadings obtained for PC1 and PC2 (Fig. 3B) might have offered information on the most important wavelengths to distinguish the cultivars, this was not useful because separation was not evident in the preceding plot (Fig. 3A). PCA maximises the variance in the first components, which may or may not be related to the segregation of the classes; this does not guarantee the class separability of data due to its unsupervised nature (Jolliffe, 2002).

3.3. Cultivar classification using individual pixel spectrum

A PLS-DA model was performed using the spectral range of 450–1040 nm and the spectrum of the individual pixels of each fruit of the calibration set. The values obtained for sensitivity and specificity (Table 2) indicated that the number of samples correctly identified as belonging to a specific cultivar, or not, was above 0.80 in the CV set, using five LV. Sensitivity of 0.83 and 0.86 was determined for 'Diamond Ray' and 'Big Top' respectively being the

accuracy of classification 84.8% and error 0.15.

Using the spatial data collected by the imaging system the combined results were applied to the calibration set. The predicted class of each pixel was obtained by introducing the spectrum measured for those pixels into the previously built model, and visualising the result. Each pixel was coloured blue if it was assigned to 'Diamond Ray' or red if it was assigned to 'Big Top', as shown in Fig. 4A. The accuracy of this classification was 83.8% and error 0.16.

To classify each fruit using this approach, the whole fruit was assigned to the class found in the majority of its pixels (Fig. 4B). In this case, the accuracy and the classification error were 84.4% and 0.16. In both cases, 'Big Top' was also the best discriminated, with a sensitivity of about 0.90.

3.4. Cultivar classification using mean fruit spectrum

The sensitivity and specificity in the results of calibration using

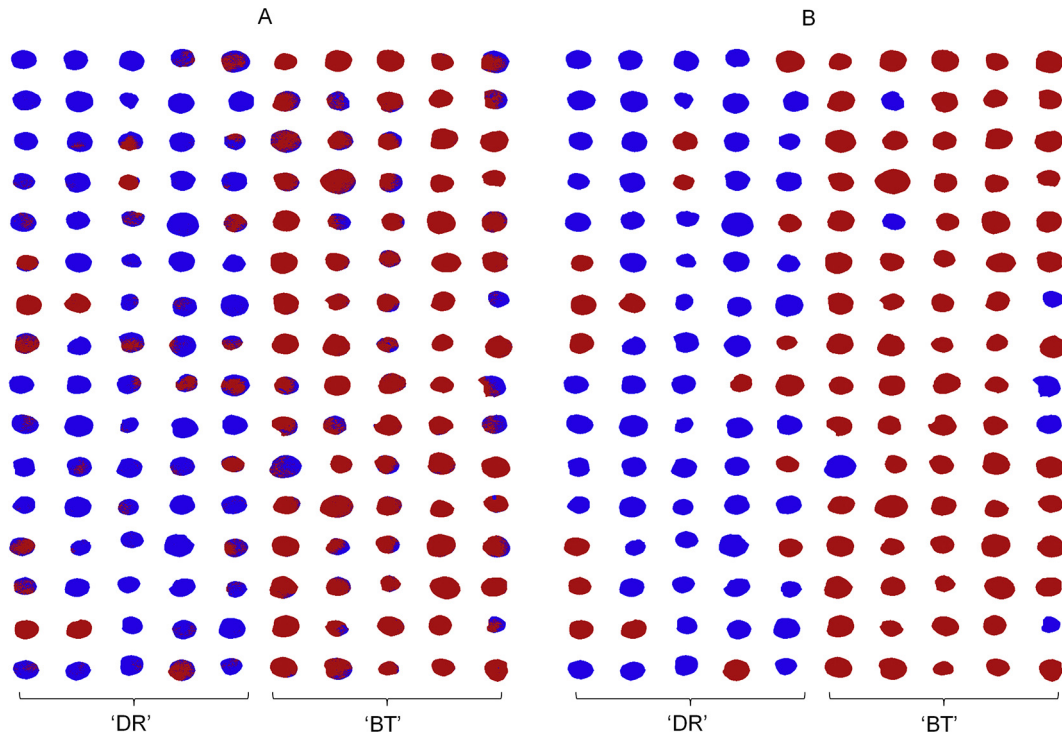


Fig. 4. Visualisation of cultivar classification using individual pixel spectrum: A) Pixel classification method; B) Object classification method.
Key for Fig. 4: Blue = 'Diamond Ray'; Red = 'Big Top'.

Table 3
Cultivar discrimination using the mean spectrum approach.

#V	#LV	Set	Class	Sensitivity	Specificity	Error	Accuracy (%)
60	6	Calibration	'DR'	0.94	0.94	0.06	93.8
			'BT'	0.94	0.94		
		Cross Validation	'DR'	0.93	0.94	0.07	93.2
			'BT'	0.94	0.93		
		Validation	'DR'	0.94	0.94	0.06	94.4
			'BT'	0.94	0.94		

V=Variables; LV = Latent variables; 'DR' = 'Diamond Ray'; 'BT' = 'Big Top'.

the mean fruit spectrum approach giving values above 0.90 using six LV (Table 3). In this case, both cultivars were discriminated similarly and the accuracy of classification of the CV was 93.2% and error 0.07.

In order to get a graphical view of the veracity of the classification obtained using the validation set, the class for each fruit was predicted by introducing the mean spectrum measured into the previously built model. The result was visualised showing the fruit coloured blue if the mean value was assigned by the model to 'Diamond Ray' or red if it was assigned to 'Big Top' (Fig. 5A). The results for the validation set were similar to those obtained in the calibration, showing an accuracy of 94.4% with a classification error of 0.06. The ANOVA results indicated that the mean spectra model was significantly better than the pixel model ($p < 0.05$) to classify the fruits.

As Williams and Kucheryavskiy (2016) pointed out, using properly computed object features as the mean spectrum decreases the amount of data, leading to more stable classification models. Furthermore, this approach avoids classifying by pixels when objects from different classes contain many similar pixels and are easily miss-assigned to the opposite class, such as for the cultivars studied in this work. On the other hand, it is important to include the negative influence of the sphericity of the fruits on the

reflectance of the light. As seen in Fig. 4A, most errors occur at the borders of the fruit, since the centres are usually well illuminated. The pixels near the borders are therefore more likely to be wrongly classified, thus affecting the overall result. In contrast when using the mean fruit spectrum, the averaging minimises these errors.

3.5. Selection of the optimal wavelengths

In order to optimise the algorithms for an automatic in-line sorting system working at high speed, it is important to reduce the computational complexity generated by the huge amount of data obtained by hyperspectral imaging systems. This problem is commonly alleviated by techniques that retain the information in the few bands that reveal the most variability and therefore most significant information in the hyperspectral image (Du & Sun, 2006). The method used in this study was the vector of the regression coefficients. This measures the association between each variable and the response and selects variables in two steps: (i) the PLS-DA model is fitted to the data, and (ii) the variable selection is based on a threshold (Mehmood, Liland, Snipen, & Sæbø, 2012). Variables with a high absolute value can be selected because they make the highest contribution to cultivar classification and those with a small absolute value can be ignored. In this study, the

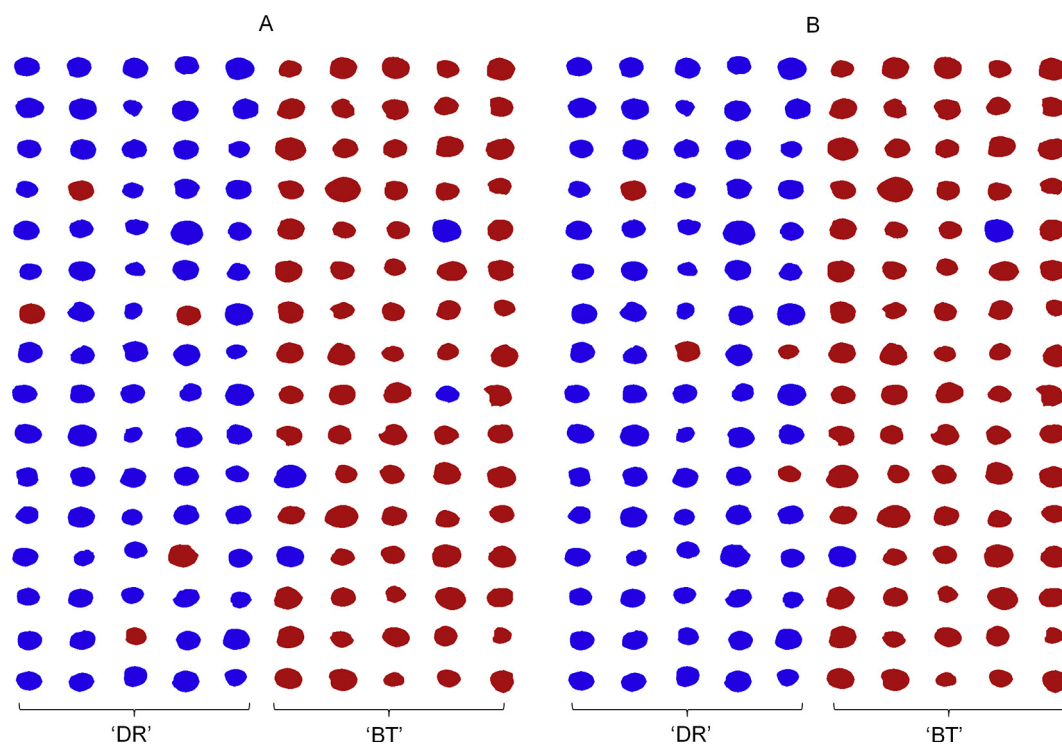


Fig. 5. Visualisation of cultivar classification using mean spectrum: A) Classification using the full range; B) Classification using 14 optimal wavelengths. Key for Fig. 5: Blue = 'Diamond Ray'; Red = 'Big Top'.

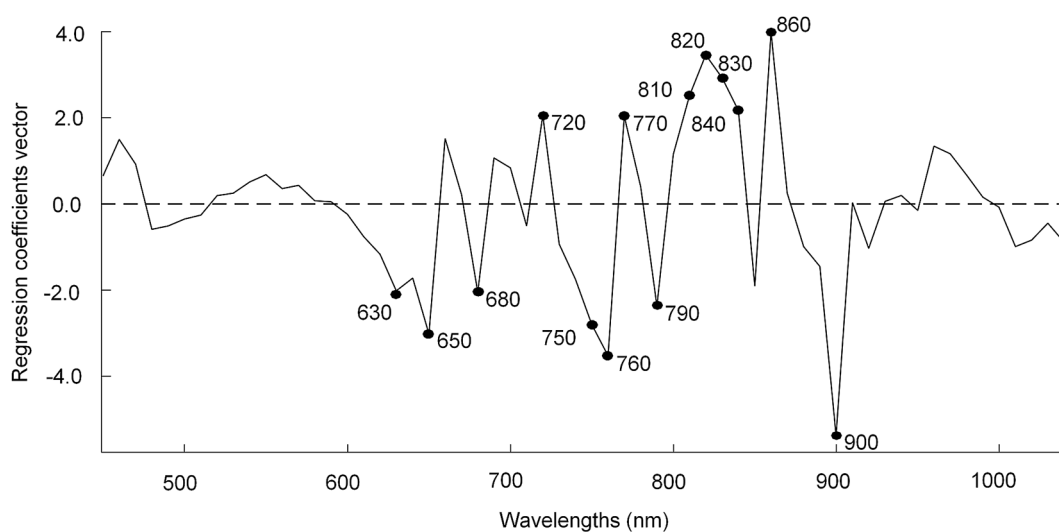


Fig. 6. Vector of regression coefficients of the PLS-DA model using mean spectra and with the optimal wavelengths selected.

Table 4

Cultivar discrimination using the mean spectrum and the optimal wavelengths methods.

#V	#LV	Set	Class	Sensitivity	Specificity	Error	Accuracy (%)
14	6	Calibration	'DR'	0.94	0.94	0.06	93.8
			'BT'	0.94	0.94		
		Cross Validation	'DR'	0.93	0.94	0.07	93.2
			'BT'	0.94	0.93		
		Validation	'DR'	0.95	0.98	0.04	96.3
			'BT'	0.98	0.95		

V=Variables; LV = Latent variables; 'DR' = Diamond Ray; 'BT' = 'Big Top'.

Table 5

Cultivar discrimination using colour imaging and by a trained panel.

	#V	#LV	Set	Class	Sensitivity	Specificity	Error	Accuracy (%)
Colour imaging (PLS-DA)	3	2	Calibration	'DR'	0.75	0.61	0.32	68.0
				'BT'	0.61	0.75		
			Cross Validation	'DR'	0.75	0.62	0.32	68.3
				'BT'	0.62	0.75		
			Validation	'DR'	0.65	0.49	0.43	56.9
				'BT'	0.49	0.65		
Trained panel	-	—	Validation	'DR'	0.54	0.55	0.46	54.5
				'BT'	0.55	0.54		

V=Variables; LV = Latent variables; 'DR' = 'Diamond Ray'; 'BT' = 'Big Top'.

regression coefficients were obtained from the PLS-DA model using the mean fruit spectrum approach, due to its higher accuracy in the classification of both cultivars.

Fig. 6 shows the vector of regression coefficients. Those peaks where the absolute value was highest were selected as important wavelengths. In the VIS region the selected wavelengths were at 630, 650, 680 and 720 nm while in the NIR region they were 750–770, 790, 810–840, 860 and 900 nm.

The optimised PLS-DA model was performed using the 14 selected wavelengths as input. The sensitivities and specificities in the CV were similar to the full model using six LV (Table 4). In the prediction set, using only the 14 wavelengths, the sensitivity for the two cultivars increased from 0.94 for both to 0.95 and 0.98, in 'Diamond Ray' and 'Big Top' respectively. Fig. 5 shows the results of both classifications, using the full spectrum (Fig. 5A) and the optimal wavelengths (Fig. 5B) in which more fruits were coloured as they should be when the wavelengths selected as the most important. However, the accuracy obtained, 96.3%, was not statistically different ($p > 0.05$) from the accuracy of the full model (96.3 and 94.4%, respectively).

3.6. Hyperspectral imaging vs. colour and visual analysis

When the validation set was classified visually by the trained panel, the same fraction of each cultivar was identified correctly (Table 5). However, the accuracy was very low, i.e. 54.5% with a classification error of 0.46. This demonstrates difficulty of the human eye to distinguish between the similar external appearances of these cultivars.

Classification by the colour data had similar accuracy (p -value > 0.05) to that achieved by the trained panel (Table 5), i.e. 56.9% accuracy and error of 0.43. This is especially poor in comparison with the results of the hyperspectral imaging using 14 wavelengths, i.e. 96.3%, error 0.04 (Table 4).

These results are in agreement with the work carried out by Nogales-Bueno, Rodríguez-Pulido, Heredia, and Hernández-Hierro (2015) that used NIR hyperspectral and colour imaging to discriminate between four red grape cultivars. Only 52% of the samples were correctly classified using colour imaging but this figure increased to 86% using hyperspectral imaging. Furthermore, Font et al. (2014) described an in-line system for verification of nectarine cultivars with close harvest times using different colour space layers of the SC histogram. The success of their technique was 100% in comparing fruits of three cultivars with a single cultivar for reference. In the same experiments, human classification achieved 86% accuracy, likely attributable to the large differences in the SC of the cultivars tested.

The high rate of accuracy in classification of these cultivars using hyperspectral imaging was important because of the external similarity of the cultivars studied. This makes it difficult to accurately identify the cultivars by colour features, although they

appear very different to consumers at the table. This is a genuine problem for the industry. Although colour imaging is a rapid and inexpensive tool, it has lower discrimination power for cultivars with very similar appearance, which necessitates the use of more VIS wavelengths and optimal wavelengths in the NIR region.

4. Conclusions

The capability of VIS-NIR hyperspectral imaging to discriminate very similar cultivars of nectarine has been demonstrated in this work.

The classification of these two cultivars by colour imaging or by a trained panel was very poor, achieving an accuracy of only 56.9% and 54.5% respectively. However, hyperspectral imaging supported by chemometric methods and optimised by reduction of the spectral and spatial information enabled classification more accurately than by traditional manual or colour-based systems, and it is also faster than destructive methods.

The use of the mean spectrum of the fruit as input of the predictive models provided classification accuracy of 94.4%. To cope with the huge amount of data captured by the hyperspectral systems, the vector of the regression coefficients of a PLS-DA model identified 14 wavelengths which were selected as optimal, producing the best classification model with a classification accuracy of 96.3%.

This technique may have potential as a tool for rapid and non-destructive cultivar discrimination, allowing the selection of fruit that is better suited to the consumer's preferences. Nevertheless, the results of this study should be confirmed on a larger sample set of fruits grown in different areas and harvested at different stages of ripeness before they can be implemented in an in-line system.

Acknowledgements

This work was partially funded by INIA and FEDER funds through project RTA2015-00078-00-00. Sandra Munera thanks INIA for the FPI-INIA grant num. 43 (CPR2014-0082), partially supported by European Union FSE funds. The authors wish to thank Fruits de Ponent (Lleida) for providing the fruit.

References

- AENOR. (1981). *Productos derivados de frutas y verduras, determinación de la acidez valorable*. UNE 34211: 1981.
- Amigo, J. M., Babamoradia, H., & Elcoroaristizabal, S. (2015). Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta*, 896, 34–51.
- Amodio, M. L., Capotorto, I., Chaudhry, M. M. A., & Colelli, G. (2017). The use of hyperspectral imaging to predict the distribution of internal constituents and to classify edible fennel heads based on the harvest time. *Computers and Electronics in Agriculture*, 134, 1–10.
- Crisosto, C. H. (1994). Stone fruit maturity indices: A descriptive. *Postharvest News and Information*, 6, 65–68.
- Crisosto, C. H., Crisosto, G. M., Echeverría, G., & Puy, J. (2006). Segregation of peach and nectarine (*Prunus persica* (L.) Batsch) cultivars according to their organoleptic characteristics. *Postharvest Biology and Technology*, 39, 10–18.

- 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 Technology*, 4, 487–504.
- Du, C. J., & Sun, D. W. (2006). Learning techniques used in computer vision for food quality evaluation: A review. *Journal of Food Engineering*, 72, 39–55.
- Echeverría, G., Cantín, C. M., Ortiz, A., López, M. L., & Graell, J. (2015). The impact of maturity, storage temperature and storage duration on sensory quality and consumer satisfaction of 'Big Top' nectarines. *Scientia Horticulturae*, 190, 179–186.
- Erkinbaev, C., Henderson, K., & Paliwal, J. (2017). Discrimination of gluten-free oats from contaminants using near infrared hyperspectral imaging technique. *Food Control*, 80, 197–203.
- Feng, C. H., Makino, Y., Oshita, S., & García-Martín, J. F. (2018). Hyperspectral imaging and multispectral imaging as the novel techniques for detecting defects in raw and processed meat Products: Current state-of-the-art research advances. *Food Control*, 165–176 (in press).
- Folch-Fortuny, A., Prats-Montalbán, J. M., Cubero, S., Blasco, J., & Ferrer, A. (2016). VIS/NIR hyperspectral imaging and N-way PLS-DA models for detection of decay lesions in citrus fruits. *Chemometrics and Intelligent Laboratory Systems*, 156, 241–248.
- Font, D., Tresanchez, M., Pallegà, T., Teixidó, M., Martínez, D., Moreno, J., et al. (2014). An image processing method for in-line nectarine variety verification based on the comparison of skin feature histogram vectors. *Computers and Electronics in Agriculture*, 102, 112–119.
- Gardner, J. L. (2007). Comparison of calibration methods for tristimulus colorimeters. *Journal of Research of the National Institute of Standards and Technology*, 112, 129–138.
- Gat, N. (2000). *Imaging spectroscopy using tunable filters: A review*. Technical report. Opto-Knowledge Systems Inc. OKSI.
- Giné-Bordonaba, J., Cantina, C. M., Larrigaudière, C., López, L., López, R., & Echeverría, G. (2014). Suitability of nectarine cultivars for minimal processing: The role of genotype, harvest season and maturity at harvest on quality and sensory attributes. *Postharvest Biology and Technology*, 93, 49–60.
- Herrero-Langreo, A., Lunadei, L., Lleó, L., Diezma, B., & Ruiz-Altisent, M. (2011). Multispectral vision for monitoring peach ripeness. *Journal of Food Science*, 2, 178–187.
- Huang, F., Zhang, S., Yang, Y., Man, Z., Zhang, X., & Wu, Y. (2015). Application of hyperspectral imaging for detection of defective features in nectarine fruit. *Transactions of the Chinese Society for Agricultural Machinery*, 11, 252–259.
- Iglesias, I., & Echeverría, G. (2009). Differential effect of cultivar and harvest date on nectarine colour, quality and consumer acceptance. *Scientia Horticulturae*, 120, 41–50.
- Iqbal, A., Sun, D.-W., & Allen, P. (2014). An overview on principle, techniques and application of hyperspectral imaging with special reference to ham quality evaluation and control. *Food Control*, 46, 242–254.
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer.
- Kandpal, L. M., Lee, S., Kim, M. S., Bae, H., & Cho, B. K. (2015). Short wave infrared (SWIR) hyperspectral imaging technique for examination of aflatoxin B1 (AFB1) on corn kernel. *Food Control*, 51, 171–176.
- Keresztes, J. C., Goodarzi, M., & Saeys, W. (2016). Real-time pixel based early apple bruise detection using short wave infrared hyperspectral imaging in combination with calibration and glare correction techniques. *Food Control*, 66, 215–226.
- Li, J., Chen, L., Huang, W., Wang, O., Zhang, B., Tian, X., et al. (2016). Multispectral detection of skin defects of bi-colored peaches based on VIS–NIR hyperspectral imaging. *Postharvest Biology and Technology*, 112, 121–133.
- López-Maestresalas, A., Keresztes, J. C., Goodarzi, M., Arazuri, S., Jaren, C., & Saeys, W. (2016). Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging. *Food Control*, 70, 229–241.
- 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 Technology*, 5, 1121–1142.
- Lu, R., & Peng, Y. (2006). Hyperspectral scattering for assessing peach fruit firmness. *Biosystems Engineering*, 93, 161–171.
- Malegori, C., Nascimento, E. J., Tonetto de Freitas, S., Pimentel, M. F., & Casiraghi, E. (2017). Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. *Talanta*, 165, 112–116.
- Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems*, 118, 62–69.
- Munera, S., Amigo, J. M., Blasco, J., Cubero, S., Talens, P., & Aleixos, N. (2017). Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging. *Journal of Food Engineering*, 214, 29–39.
- Munera, S., Besada, C., Aleixos, N., Talens, P., Salvador, A., Sun, D.-W., et al. (2017). Non-destructive assessment of the internal quality of intact persimmon using colour and VIS/NIR hyperspectral imaging. *LWT - Food Science and Technology*, 77C, 241–248.
- Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., et al. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46, 99–118.
- Nogales-Bueno, J., Rodríguez-Pulido, F. J., Heredia, F. J., & Hernández-Hierro, J. M. (2015). Comparative study on the use of anthocyanin profile, color image analysis and near-infrared hyperspectral imaging as tools to discriminate between four autochthonous red grape cultivars from La Rioja (Spain). *Talanta*, 131, 412–416.
- Pan, L., Zhang, Q., Zhang, W., Sun, Y., Hua, P., & Tu, K. (2016). Detection of cold injury in peaches by hyperspectral reflectance imaging and artificial neural network. *Food Chemistry*, 192, 134–141.
- Pérez-Marín, D., Sánchez, M. T., Paz, P., González-Dugo, V., & Soriano, M. A. (2011). Postharvest shelf-life discrimination of nectarines produced under different irrigation strategies using NIR-spectroscopy. *LWT - Food Science and Technology*, 44, 1405–1414.
- Rajkumar, P., Wang, N., Elmasry, G., Raghavan, G. S. V., & Gariepy, Y. (2012). Studies on banana fruit quality and maturity stages using hyperspectral imaging. *Journal of Food Engineering*, 108, 194–200.
- Reig, G., Alegre, S., Gatiús, F., & Iglesias, I. (2013). Agronomical performance under Mediterranean climatic conditions among peach [*Prunus persica* (L.) Batsch] cultivars originated from different breeding programs. *Scientia Horticulturae*, 150, 267–277.
- Reig, G., Iglesias, I., & Echeverría, G. (2009). Agronomical performance, fruit quality and sensory attributes of several flat peach and flat nectarine cultivars. In *VII international peach symposium* (Vol. 962, pp. 563–569).
- Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28, 1201–1222.
- Serranti, S., Cesare, D., Marini, F., & Bonifazi, G. (2013). Classification of oat and groat kernels using NIR hyperspectral imaging. *Talanta*, 103, 276–284.
- Sun, Y., Wang, Y., Xiao, Y., Gu, X., Pan, L., & Tu, K. (2017). Hyperspectral imaging detection of decayed honey peaches based on their chlorophyll content. *Food Chemistry*, 235, 194–202.
- USDA. (2016). *EU-28 stone fruit annual*. <https://gain.fas.usda.gov/> Accessed 15.04.17.
- Valero, C., Crisosto, C. H., & Slaughter, D. (2007). Relationship between nondestructive firmness measurements and commercially important ripening fruit stages for peaches, nectarines and plums. *Postharvest Biology and Technology*, 44, 248–253.
- Vélez-Rivera, N., Gómez-Sanchis, J., Chanona-Pérez, J. J., Carrasco, J. J., Millán-Giraldo, M., Lorente, D., et al. (2014). Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning. *Biosystems Engineering*, 122, 91–98.
- Verdú, S., Vázquez, F., Grau, R., Ivorra, E., Sanchez, A. J., & Barat, J. M. (2016). Detection of adulterations with different grains in wheat products based on the hyperspectral image technique: The specific cases of flour and bread. *Food Control*, 62, 373–380.
- Williams, P. J., & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, 209, 131–138.
- Yang, C. H., Sun, D. W., Pu, H., Wang, N. N., & Zhu, Z. (2015). Rapid detection of anthocyanin content in lychee pericarp during storage using hyperspectral imaging coupled with model fusion. *Postharvest Biology and Technology*, 103, 55–65.
- Zhang, B., Li, J., Fan, S., Huang, W., Zhao, C., Liu, C., et al. (2015). Hyperspectral imaging combined with multivariate analysis and band math for detection of common defects on peaches (*Prunus persica*). *Computers and Electronics in Agriculture*, 114, 14–24.
- Zhu, N., Lin, M., Nie, Y., Wu, D., & Chen, K. (2016). Study on the quantitative measurement of firmness distribution maps at the pixel level inside peach pulp. *Computers and Electronics in Agriculture*, 130, 48–56.