In-line application of visible and near-infrared diffuse reflectance spectroscopy to identify apple varieties

Cortés, V.[1]; Cubero, S.[2]; Blasco, J.[2]; Aleixos, N.[3]; Talens, P.[1*]

[1]Departamento de Tecnología de Alimentos. Universitat Politècnica de València. Camino de Vera s/n, 46022, Valencia (Spain).

[2]Centro de Agroingeniería. Instituto Valenciano de Investigaciones Agrarias (IVIA). Ctra. CV-315, km. 10,7, 46113, Moncada, Valencia (Spain).

[3]Departamento de Ingeniería Gráfica. Universitat Politècnica de València. Camino de Vera s/n, 46022, Valencia (Spain).

*Corresponding author: Pau Talens. Departamento de Tecnología de Alimentos. Universitat Politècnica de València. Camino de Vera, s/n. 46022. Valencia. Spain. Phone: 34-963879836, e-mail: pautalens@tal.upv.es

ABSTRACT

One of the most studied techniques for the non-destructive determination of the internal quality of fruits has been visible and near-infrared (VIS-NIR) reflectance spectroscopy. In the case of apples, it has been applied mostly for internal quality assessment and variety distinction. However, this technique has traditionally been used in laboratory setups for individual fruit inspection. This work evaluates a new non-destructive in-line VIS-NIR spectroscopy prototype to identify five apple varieties, with the advantage that it allows the spectra to be captured with the probe at the same distance from all the fruits regardless of their size. The prototype was tested using varieties with a similar appearance by acquiring the diffuse reflectance spectrum of the samples travelling at a speed of 1 fruit/s. Principal component analysis (PCA) was used to determine the

1

variables that explain the most variance in the spectra. Seven principal components were then used to perform linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). QDA was found to be the best in-line classification method, achieving 98% and 85% success rates for red and yellow apple varieties, respectively. The results indicated that the in-line application of VIS-NIR spectroscopy that was developed is potentially feasible for the detection of apple varieties with an accuracy that is similar to or better than a laboratory system.

1. INTRODUCTION

More than 7500 apple (*malus sp., Rosaceae*) varieties are currently cultivated (Wu et al. 2017) with significant differences in qualities, which makes this one of the most consumed fruits in the world (Ronald & Evans 2016). Consumers have different tastes and preferences from one country or region to another. However, there are trends that are common to all of them, such as an increasing demand for higher quality, both external and internal. The external properties (presentation, appearance, uniformity, maturity, freshness, etc.) are the main reason governing the purchase decision of fruits, which is normally taken when the consumer sees the products exhibited on the shelves. The internal quality (flavour, aroma, texture, firmness, acidity, sugar content, absence of contaminants, etc.) (Wojdyło et al. 2008) is linked to aspects that are generally not perceptible at first sight, although future purchasing decisions may depend on them. Indeed, because of the variety of consumers, it is difficult to establish a specific definition of quality, but generally a fruit has better quality when it is superior in one or several

attributes that meet consumer expectations (Kader et al. 1985). However, no matter how much effort is put into obtaining products with excellent quality, this consumer satisfaction will be defrauded when, either by mistake or by fraud, fruits, like apples, of similar appearance (external quality) but different internal quality are mixed in the market place (López 2003).

Generally, several varieties of apples can be grown simultaneously in a single apple orchard, and they can be sold at the same time. Therefore, different apple varieties can easily be mixed during harvesting and marketing when they have a similar appearance. Proper identification of fruits has to be done by workers in a subjective and manual way because physical or chemical analytical methods for apple evaluation are expensive and time-consuming. This sorting results in high costs, tediousness and inconsistency associated with human beings (Ronald & Evans 2016). Therefore, apple sellers need new electronic non-destructive techniques to distinguish apple varieties (Shang et al. 2015). Several non-destructive techniques to discriminate among apple varieties have already been attempted. Visible and near-infrared (VIS-NIR) spectroscopy has been used by different researchers such as He et al. (2007), Wu et al. (2016), Luo et al. (2011), Wu et al. (2017) and Song et al. (2016). All of these studies only extract information from the VIS-NIR spectra, or even only from the NIR spectra, of the apples for off-line discrimination among varieties; however, none of them develop in-line applications to classify apple varieties. Such an application is developed in this study through a preliminary in-line prototype. Other non-destructive techniques used to classify apple varieties have been dielectric properties, solid-state sensors (ISFET), digital image analysis or electronic nose. Shang et al. (2015) used 160 apples of three varieties in an attempt to discriminate them according to the dielectric spectra measured at laboratory scale. Alonso et al. (2003) analysed the ISFET by ions such as calcium, potassium and

nitrates, both in apple juice and in in situ apple fruit. Regarding digital image analysis, Sabanci and Ünlersen (2016) and Ronald and Evans (2016) captured apple images and obtained a good classification using the measurement of different properties (colour and size). Finally, an electronic nose was developed by Marrazzo et al. (2005), who tested the feasibility of detecting the difference among volatile gases emitted from different intact apple varieties, but good results in terms of the classification of the three varieties was only obtained for day 1. Among these non-destructive methods, the most widely used is VIS-NIR spectroscopy, because it is extremely fast, chemical-free and suitable for in-line use (Wu et al. 2017). The literature contains different studies on the applicability of in-line VIS-NIR spectroscopy. For example, Jie et al. (2014) developed an in-line prototype based on the VIS-NIR technology in the range of 687-920 nm for the prediction of the soluble solid content (SSC) of watermelon and achieved r_{pre} of 0.70 and RMSEP of 0.33 ºBrix. Brunt et al. (2010) measured the chemical composition of potatoes by an automated semi-industrial system with in-line NIR reflectance. The system was only able to process 12 potato samples per hour, but the starch and the coagulating protein concentrations showed a good NIR prediction. Sun et al. (2016) determined the SSC and identified brown core pears simultaneously in-line by VIS-NIR transmittance spectroscopy. The spectra between 600-904 nm were recorded at a moving speed of five samples per second. The classification accuracy of brown core pears was 98.3% and the predictive precision of SSC was around 98%. Similarly, Shenderey et al. (2010) evaluated the ability of VIS-NIR mini-spectrometers (400-1000 nm) to detect mouldy core in 'Red Delicious' apples using an in-line development. The accuracy of the classification was higher than 92% but, in this case, although the apples were in movement, both the probe and the apples were always at fixed positions during the measurements. Salguero-Chaparro et al. (2012) evaluated the effect of some parameters, such as focal distance and integration time, on

the spectral repeatability for the analysis of intact olive fruits on a conveyor belt in the spectral range of 380-1690 nm using a diode array spectrometer. Regarding this last research, it should be pointed out that one of the major drawbacks hindering the incorporation of technology based on spectroscopy into automated in-line sorters is the large differences in size and shape that can be found from one fruit to another. Since the probe is located at a fixed point, the distance between the probe and the fruit will vary depending on the size of the fruit. However, it is important to ensure that the distance between the probe and the fruit is close enough to obtain a good spectrum, but not so near that it saturates the signal, and this distance must be constant for all objects. If the fruit is very close, the signal saturates even with very low integration times. This is because light practically does not penetrate inside the object and is only reflected on its surface towards the detector. Conversely, a distance of more than 13 mm (Salguero-Chaparro et al. 2012) causes the detector to receive hardly any of the light reflected on the object. In addition, the optical configuration, the intensity of the light and the distance between the probe and the fruit are parameters that greatly influence the depth of penetration of light (Lammertyn et al. 2000). Therefore, to avoid a negative influence on the measurements, it is very important to develop mechanisms to ensure that the distance between the fruit and the probe remains stable for any fruit in the inspection line, regardless of its size or shape. Since there is no previous research about the in-line application of VIS-NIR diffuse reflectance spectroscopy to classify apple varieties, the objectives of this work are (1) to develop a prototype system, using VIS-NIR reflectance spectroscopy, to be used for the in-line non-destructive measurement of apples, (2) to use an automated system that ensures that the distance between the probe and the fruit is the same regardless of the size of the fruit, and (3) to use the data obtained by the in-line system to differentiate apple varieties using chemometric methods.

2. MATERIALS AND METHODS

2.1. Apple samples

With the aim of carrying out the study using fruits with different internal and external properties, a total of 500 apples of five different varieties and origins were purchased from a local market (Figure 1). The set of fruits used in the experiments consisted of 100 pieces of each of the following varieties: Fuji (FU) from the region of Aquitaine (France), Red Delicious (RD) and Royal Gala (GA) from L'Alt Empordà (Spain), Golden Delicious (GD) from the Protected Geographical Indication of 'Alto Adige' (Italy) and Golden Delicious from the Protected Designation of Origin 'Pomme du Limousin' (France), known as Golden Rosé (GR). The reason for using Golden Delicious apples coming from two different origins was to know whether the system could also recognise fruits of the same variety but cultivated under different conditions and therefore with different properties.

The average diameter of each apple was measured by using a digital calliper in the equatorial and polar sections. The weight of each apple was measured using an electronic scale (Entris2202-1S, Sartorius Lab Instruments GmbH & Co. KG, Goettingen, Germany). The characteristic features of the fruits used in this study are summarised in Table 1.

2.2. VIS-NIR instrumentation

Diffuse reflectance of the apples was measured using a multichannel spectrometer platform (AvaSpecAS-5216 USB2-DT, Avantes BV, The Netherlands) equipped with two detectors. The first detector (AvaSpec-ULS2048 StarLine, Avantes BV, The Netherlands) included a 2048-pixel charge-coupled device (CCD) sensor (SONY

ILX554, SONY Corp., Japan), 50 μm entrance slit and a 600-line mm-1 diffraction grating covering the visible and near-infrared range from 600 nm to 1100 nm (VNIR) with a spectral FWHM (full width at half maximum) resolution of 1.15 nm and a spectral sampling interval of 0.255 nm. The second detector (AvaSpec-NIR256-1.7 NIRLine, Avantes BV, The Netherlands) was equipped with a 256-pixel non-cooled InGaAs (Indium Gallium Arsenide) sensor (Hamamatsu 92xx, Hamamatsu Photonics K.K., Japan), 100 μm entrance slit and a 200-line mm-1 diffraction grating covering the near-infrared range from 900 nm to 1700 nm (NIR) with a spectral FWHM resolution of 12 nm and a spectral sampling interval of 3.535 nm.

The measurements were performed using a bi-directional fibre-optic reflectance probe (FCR-7IR200-2-45-ME, Avantes BV, The Netherlands). The probe is fitted with an illumination leg, composed of six 200 μm diameter fibres, which connects to a stabilised 10 W tungsten halogen light source (AvaLight-HAL-S, Avantes B0V, The Netherlands) and the other leg of the fibre-optic probe, with a 200 μm diameter fibre, was connected to both detectors for simultaneous measurement using another Y-shaped fibre-optic probe (FCB-IR200-2-ME, Avantes BV, The Netherlands). The probe tip was designed to provide reflectance measurements at an angle of 45° so as to minimise specular reflectance from the surface of the fruit. To control the detectors and to acquire the spectra, a personal computer equipped with a commercial software package (AvaSoft version 7.2, Avantes, Inc.) was used.

A 99% reflective white tile (WS-2, Avantes BV, The Netherlands) was used to calibrate the system so that the maximum reflectance value over each range of wavelengths was around 90% of saturation (Lorente et al. 2015).

2.3. VIS-NIR in-line system and spectra acquisitions

All the samples were equilibrated to room temperature (25 °C) before obtaining the spectral measurements. During the in-line test, the fruits were transported continuously using a conveyor belt composed of bi-conical rollers while they were being measured. Fruits were placed manually on the conveyor belt with the stem at the top to capture measurements of the equatorial part. To achieve a uniform distance between the probe and the fruits, a programmable electronic device (Raspberry Pi3, Raspberry Pi Foundation, United Kingdom) together with a small high-resolution IR camera (Raspberry Pi NIR Camera v2, Raspberry Pi Foundation, United Kingdom) were placed at a height of 25 cm above the conveyor belt and exactly perpendicular to the centre of the rollers. Illumination of the scene was achieved using two IR LED-based light sources synchronised with the camera module. A computer vision application captured an image of each fruit, detecting its exact location and estimating the diameter across the width of the rollers. The distance between the centre of the rollers and the end of the fruit on the measuring side are used to enable the probe to approach the fruit. Figure 2 shows all the parameters and components of the system.

To synchronise the acquisition of images and the movement of the probe with the advance of the apples, an optical encoder (WDG 40S-500-ABN-G05-K1, Wachendorff Elektronik, Germany) was coupled to the traction roller of the machine.

Using the distance between the fruit and the probe estimated by the image analysis system, and the position of the fruit on the conveyor belt estimated from the pulses given by the encoder, the electronic device activated a linear actuator composed of a stepper motor and an endless screw to bring the probe accurately towards the fruit. To calculate the displacement of the probe (d0), the distance between the centre of the rollers and the resting position of the probe (d1), the known distance between the centre of the rollers and the perimeter of the fruit on the measuring side given by the image analysis system

(d2), and the distance between the probe and the fruit (d3) were used according to the distances shown in Figure 2 and Equation 1. All values were calculated in millimetres.

$$d0 = d1 - d2 - d3 \qquad (1)$$

The process of bringing the probe closer to the fruit is shown in Figure 3. Different integration times were tested in preliminary tests at different distances between the fruits and the probe, with the aim of setting the minimum integration time that gives a good signal for the in-line application. Subsequently, the integration time was set to 150 ms for the VNIR detector and to 500 ms for the NIR detector. Each fruit was passed twice in the line to obtain two measurements of the diffuse reflectance spectrum. Due to the limitation of the integration time of the NIR detector, the speed of the conveyor belt had to be limited to 0.81 m/s.

2.4. Data analysis

2.4.1. Chemometric analyses

The chemometric analyses of the collected spectra were performed using the Unscrambler V10.4 software package (CAMO ASA, Oslo, Norway). A matrix, where the rows represent the number of samples (N = 500 samples) and the columns represent the variables (X-variables and Y-variable), was produced with the spectral data. The spectral signals from the two detectors were the predictors, or X-variables. The categorical 'dummy' variable created by assigning different letters to the different apple varieties was the response, or the Y-variable.

The raw spectra were transformed to apparent absorbance (log (1/R)) values to linearise the correlation with the concentration of the constituents (Hernández et al. 2006;

Shao et al. 2007; Liu et al. 2009). To prevent a low signal-noise ratio occurring at the limits of the spectral sensitivity of the equipment used, only wavelengths ranging from 600 to 1700 nm were included in this study.

Because the high resolution captured in the VNIR range introduced increased noise in the signal (Cortés et al. 2016), the VNIR spectra were pre-processed using a reduction factor of 10. Savitzky-Golay smoothing (the segment size is 3) was applied to reduce the effects caused by the physiological variability of samples (Carr et al. 2005; Beghi et al. 2017). Due to scattering effects, the light does not always travel the same distance inside the sample before it is collected by the probe (Santos et al. 2013). A longer path travelled by the light results in a lower relative reflectance value, since more light is absorbed. This effect causes a parallel translation of the spectra that needs to be corrected to avoid a negative influence in the calibration models. A typical method to perform this correction, and the one used in this work, is the EMSC technique (He et al. 2007; Martens et al. 2003; Bruun et al. 2007). Those two pre-treatments were considered the best results for the VNIR spectra, and those two pre-processing methods and the second derivate with Gap-Segment (2.3) were the best results for the NIR spectra (Cortés et al. 2016).

After the pre-processing, principal component analysis (PCA) (Næs et al. 2002) was applied to explore a possible clustering of the sample spectra that could be associated to the different apple varieties. This technique is a linear and unsupervised procedure that allows useful information to be extracted from the data and to explore the data structure and the relationship between objects as well as the global correlation of the variables (Beebe et al. 1998).

In addition, the linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) classification methods were applied on the spectra of the apples in order to classify the samples as belonging to one of the five varieties. The objective of these

methods is to find models that allow the maximum separation among different classes in a set of objects. The difference between LDA and QDA is that LDA uses pooled covariance to assign an unknown sample to one of the pre-defined groups, while QDA uses the covariance of each of the groups instead of pooling them (Næs et al. 2002). In both cases, the number of samples in the training set must be larger than the number of variables included in the model (Kozak & Scaman 2008; Sádecká et al. 2016), thus a reduction of the dimensionality is necessary. This was achieved by using the PCA scores as input data in the model instead of the raw variables, since the linear combinations of the original variables given by the PCA, called principal components (PCs), are uncorrelated (Rodríguez-Campos et al. 2011).

To develop the discriminant model, a training set composed of 80% of the samples selected at random was used. The model was internally validated using full cross-validation (CV; leave-one-out method) (Casale et al. 2008; Huang et al. 2008). Predictions were carried out using the evaluation set composed of the remaining 20% of the samples (Soares et al. 2013). The performance of the model was evaluated by accuracy, which is defined using the ratio of samples in the test set correctly assigned to their respective classes presented using confusion matrices.

3. RESULTS AND DISCUSSION

3.1. Features of VIS-NIR Spectra

Preliminary tests were carried out on both red and yellow apples to determine the correct measuring distance between the probe and the fruit for each type of fruit. To do so, distances of 1, 5 and 10 mm were tested with the fruits travelling on the conveyor belt

at the same speed (one fruit per second). The average spectra of yellow and red apples captured using different distances between the light source/detection probe and the fruit are shown in Figure 4. When the probe was very close to the fruit (1 mm), it was observed that the signal saturated in some samples even with very low integration times. This is because practically no light penetrates inside the object and is only reflected on its surface towards the detector. Conversely, using a distance of 10 mm, the detector receives very little reflection, which is in accordance with Salguero-Chaparro et al. (2012). Therefore, for this study, the diffuse reflectance spectrum of all fruits was acquired at 5 mm from the fruit, regardless of their size. In Equation 1, parameter d3 has a value of 5 mm.

Once the measurement distance had been determined, VIS-NIR spectra were measured for all apples. Figure 5 shows the raw spectra of all samples. The differences in colour among the apple varieties were detected in the VIS wavelength range (i.e. 600-700 nm). However, the acquired spectra exhibit similar profiles across the spectral range that was studied, and even overlapped heavily, especially the spectra between 730-1700 nm. According to Fernández-Ahumada et al. (2006), the use of the VIS and NIR wavelength range offers great expectations for NIR characterisation in some products such as potatoes. Because it was difficult to discriminate the five kinds of apples just from the overlapping spectra, it was necessary to use feature extraction methods, such as PCA, LDA and QDA, to distinguish the apple varieties.

3.2. Clustering analyses based on PCA

The PCA model, as previously mentioned, was first constructed in order to explore a possible clustering of the spectra of the samples that could be associated to the different apple varieties. Figure 6 shows the training spectra set plotted according to the values of the PCA scores. Each variety is shown with a different colour and shape for a better

visualisation. As shown in this figure, red and yellow varieties can be separated easily due to the different spectral reflectance represented in the visible spectrum. However, colour and appearance are not always useful for discriminating apple varieties. There are some varieties that are very difficult to differentiate visually on the production line but show important compositional differences that are represented in the NIR spectrum, thereby affecting the taste and consequently their acceptance by the final consumers.

Additionally, the PCA model was used to reduce the original variables for constructing the LDA and QDA models. The cumulated reliability of the seven PCs reaches 95.57% (Table 2). Hence the first seven PCs were adopted as the inputs in the LDA and QDA models.

3.3. Classification by LDA and QDA models

PCA can analyse the varieties of apples qualitatively, but it cannot discriminate these varieties quantitatively. In this study, LDA and QDA were used to build a quantitative analysis model to discriminate apple varieties, shown below. Table 3 provides the confusion matrices that summarise the classification results using LDA and QDA techniques. From these matrices, it can be concluded that no red apple was classified as a yellow variety and vice versa. This expected correct classification is mainly due to the contribution of the visible spectrum mentioned previously. In general, QDA achieved better results than LDA for both types of samples, probably because LDA can only learn linear boundaries, while QDA can learn quadratic boundaries, LDA being a much less flexible classifier than QDA (James et al. 2014). The results show that classification accuracy for LDA model was 82.5% and for the QDA model it was 85% for the same test set of yellow varieties that was used. Similarly, the red varieties obtained a classification accuracy for the LDA model of 75% and of 98.3% for the QDA model.

The differences were observed mainly in the high classification accuracy achieved by the QDA model in red apples. However, it is important to note that the red apples were of different varieties, while the yellow ones both belonged to the same variety but different origins, thus demonstrating that this method can not only discriminate between varieties but also by the origin of the apple.

He et al. (2007) already used PCA and an artificial neural network (ANN) coupled with VIS-NIR (400-960 nm) to classify three red apple varieties ('Fuji' from China, 'Red Delicious' from USA and 'Copefrut Royal Gala' from USA). They used only 30 apples of each variety, obtaining an accuracy of 100% but using a static arrangement. Wu et al. (2016) applied four kinds of feature extraction methods, namely PCA, PCA+LDA, discriminant partial least squares (DPLS) and sorting discriminant analysis (SDA), on the NIR spectra (1000-2500 nm) of two varieties of apples with a red external colour ('Huaniu' and 'Fuji' from China). In their research, the SDA model was the best one with a classification accuracy of 96.67% for discriminating apple varieties. In these previous studies, He et al. (2007) and Wu et al. (2016) only classified red apple varieties on a laboratory scale using static measurements. In contrast, the new in-line detection prototype has been developed for spectral analysis while the fruit is transported by a conveyor belt in the processing line. The in-line system developed here allowed us to obtain as good or better results than previous studies involving laboratory systems. The speed of advance of the fruits when the spectra were measured was one fruit per second. This was the same speed as that used by Shenderey et al. (2010) in their experiments when they measured the spectra of apples in movement to detect those affected by mouldy core. In their work, the measuring device was placed on a rotating table that served as a laboratory device to test the spectrometry in moving fruits, thus demonstrating that it is possible to perform spectrometry on apples in-line. However, in this work, the practical

implementation was performed in an industrial prototype with what was similar to a commercial conveyor belt and presents certain novel advantages such as ensuring the same measuring distance for all fruits.

4. CONCLUSIONS

A novel prototype for the in-line identification of apple varieties based on VIS-NIR spectroscopy has been developed and tested. One of the main disadvantages of the in-line system when fruits are travelling on a conveyor belt is that the diversity of fruit sizes causes the spectra to be taken at different distances, since the position of the probe is fixed. This has a negative effect on the robustness of the results obtained. This work proposes a new device that allows the measurements to be captured at always the same distance from the fruit, thus ensuring that they are all taken in the same conditions. Two classification methods were used to determine the variety of the apples and, in the case of Golden Delicious, to detect fruits of different origins. QDA yielded better results than LDA reaching a 98% and 85% rate of success for red and yellow varieties, respectively. The proposed prototype may contribute to improve the control of production as a tool for detecting fraud in a non-destructive and highly accurate manner. However, in order to effectively transfer this system to industry, the speed of one fruit per second needs to be improved.

5. ACKNOWLEDGEMENTS

REFERENCES

Alonso, J., Artigas, J., & Jimenez, C. (2003). Analysis and Identification of Several Apple Varieties Using ISFETs Sensors. Talanta, 59, 1245-1252.

Beebe, K.R., Pell, R.J., & Seasholtz, M.B. (1998). In: Chemometrics: a Practical Guide, New York, USA. John Wiley and Sons.

Beghi, R., Giovenzana, V., Brancadoro, L., & Guidetti, R. (2017). Rapid evaluation of grape phytosanitary status directly at the check point station entering the winery by using visible/near infrared spectroscopy. Journal of Food Engineering, 204, 46-54.

Brunt, K., Smits, B., & Holthuis, H. (2010). Design, construction, and testing of an automated NIR in-line analysis system for potatoes. Part II. Development and testin of the automated semi-industrial system with in-line NIR for the characterization of potatoes. Potato Research, 53, 41-60.

Bruun, S.W., Sondergaard, I., & Jacobsen, S. (2007). Analysis of protein structures and interactions in complex food by near-infrared spectroscopy. 1. Gluten powder. Journal of Agricultural and Food Chemistry, 55, 7234-7243.

Carr, G.L., Chubar, O., & Dumas, P. (2005). Spectrochemical Analysis Using Infrared Multichannel Detectors. In Bhargava, R., Levin, I.W. (Eds.), 1st ed. Wiley-Blackwell, Oxford, 56-84.

Casale, M., Casolino, C., Ferrari, G., & Forina, M. (2008). Near infrared spectroscopy and class modelling techniques for geographical authentication of Ligurian extra virgin olive oil. Journal of Near Infrared Spectroscopy, 16, 39-47.

Cortés, V., Ortiz, C., Aleixos, N., Blasco, J., Cubero, S., & Talens, P. (2016). A new internal quality index for mango and its prediction by external visible and near infrared reflection spectroscopy. Postharvest Biology and Technology, 118, 148-158.

Fernández-Ahumada, E., Garrido-Varo, A., Guerrero-Ginel, A.E., Wubbels, A., van der Sluis, C., & van der Meer, J.M. (2006). Understanding factors affecting near infrared analysis of potato constituents. Journal of Near Infrared Spectroscopy, 14, 27-35.

He, Y., Li, X., & Shao, Y. (2007). Fast discrimination of apple varieties using Vis/NIR spectroscopy. International Journal of Food Properties, 10(1), 9-18.

Hernández, A., He, Y., & García, A. (2006). Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. Journal of Food Engineering, 77, 313-319.

Huang, H., Yu, H., Xu, H., & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. Journal of Food Engineering, 87(3), 303-313.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). An Introduction to Statistical Learning: With Applications in R. New York: springer.

Jie, D., Xie, L., Rao, X., & Ying, Y. (2014). Using visible and near infrared diffuse transmittance technique to predict soluble solids content of watermelon in an on-line detection system. Postharvest Biology and Technology, 90, 1-6.

Kader, A.A., Kasmire, R.F., Mitchell, F.G., Reid, M.S., Sommer, N.F., & Thompson, J.F. (1985). Postharvest technology of horticultural crops. Cooperative Extension, University of California. Special Publication, 3311, 192.

Kozak, M., & Scaman, C. H. (2008). Unsupervised classification methods in food sciences: discussion and outlook. Journal of the Science of Food and Agriculture, 88, 1115-1127.

Lammertyn, J., De Baerdemaeker, J., & Nicolaï, B. (2000). Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. Postharvest Biology and Technology, 18(2), 121-132.

Liu, F., Jiang, Y., & He, Y. (2009). Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: a case study to determine soluble solids content of beer. Anal. Chim. Acta 635, 45-52.

López, A.F. (2003). 'Manual para la preparación y venta de frutas y hortalizas, del campo al mercado'. PDF File: Boletín de servicios agrícolas de la FAO, 151. http://www.fao.org/tempref/docrep/fao/006/y4893S/y4893S00.pdf

Lorente, D., Escandell-Montero, P., Cubero, S., Gómez-Sanchis, J., & Blasco, J. (2015). Visible-NIR reflectance spectroscopy and manifold learning methods applied to the detection of fungal infections on citrus fruit. Journal of Food Engineering, 163, 17-21.

Luo, W., Huan, S., Fu, H., Wen, G., Cheng, H., Zhou, J., Wu, H., Shen, G., & Yu, R. (2011). Preliminary study on the application of near infrared spectroscopy and pattern recognition methods to classify different types of apples. Food Chemistry, 128, 555-561.

Marrazzo, W.N., Heinemann, P.H., Crassweller, R.E., & LeBlanc, E. (2005). Electronic Nose Chemical Sensor Feasibility Study for the Differentiation of Apple Cultivars. American Society of Agricultural Engineers, 48 (5), 1995-2002.

Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. Analytical Chemistry, 75, 394-404.

Næs, T., Isaksson, T., Fearn, T., & Davies, T. (2002). A user-friendly guide to multivariate calibration and classification. Chichester: NIR Publications.

Rodríguez-Campos, J., Escalona-Buendía, H. B., Orozco-Avila, I., Lugo-Cervantes, E., & Jaramillo-Flores, M. E. (2011). Dynamics of volatile and non-volatile compounds

in cocoa (Theobroma cacao L.) during fermentation and drying processes using principal components analysis. Food Research International, 44, 250-258.

Ronald, M., & Evans, M. (2016). Classification of selected apple fruit varieties using Naive Bayes. Indian Journal of Computer Science and Engineering, 7(1), 13-19.

Sabanci, K., & Ünlersen, M.F. (2016). Different apple varieties classification using kNN and MLP algorithms. International Journal of Intelligent Systems and Applications in Engineering, 4, 166-169.

Sádecká, J., Jakubíková, M., Májek, P., & Kleinová, A. (2016). Classification of plum spirit drinks by synchronous fluorescence spectroscopy. Food Chemistry, 196, 783-790.

Salguero-Chaparro, L., Baeten, V., Abbas, O., & Peña-Rodríguez, F. (2012). On-line analysis of intact olive fruits by vis-NIR spectroscopy: optimisation of the acquisition parameters. Journal of Food Engineering, 112, 152-157.

Santos, P., Santos, F., Santos, J., & Bezerra, H. (2013). Application of extended multiplicative signal correction to short-wavelength near infrared spectra of moisture in marzipan. Journal of Data Analysis and Information Processing, 1, 30-34.

Shang, L., Guo, W., & Nelson, S. O. (2015). Apple variety identification based on dielectric spectra and chemometric methods. Food Anal. Methods, 8, 1042-1052.

Shao, Y., He, Y., Gómez, A.H., Pereir, A.G., Qiu, Z. & Zhang, Y. (2007). Visible/near infrared spectrometric technique for nondestructive assessment of tomato 'Heatwave' (Lycopersicumesculentum) quality characteristics. Journal of Food Engineering, 81 (4), 672-678.

Shenderey, C., Shmulevich, I., Alchanatis, V., Egozi, H., Hoffman, A., Ostrovsky, V., Lurie, S., Arie, R.B., & Schmilovitch, Z. (2010). NIRS detection of moldy core in apples. Food Bioprocess Technology, 3, 79-86.

Soares, S. F. C., Gomes, A. A., Galvão Filho, A. R., Araújo, M. C. U., & Galvão, R. K. H. (2013). The successive projections algorithm. Trends in Analytical Chemistry, 42, 84-98.

Song, W., Wang, H., Maguire, P., & Nibouche, O. (2016). Differentiation of organic and non-organic apples using near infrared reflectance spectroscopy – A pattern recognition approach. Sensors, 1-3.

Sun, X., Liu, Y., Li, Y., Wu, M., & Zhu, D. (2016). Simultaneous measurements of Brown core and soluble solids content in pear by on-line visible and near infrared spectroscopy. Postharvest Biology and Technology, 116, 80-87.

Wojdyło, A., Oszmiański, J., & Laskowski, P. (2008). Polyphenolic compounds and antioxidant activity of new and old apple varieties. Journal of Agricultural and Food Chemistry, 56(15), 6520-6530.

Wu, X., Wu, B., Sun, J., & Yang, N. (2017). Classification of Apple varieties using near infrared reflectance spectroscopy and fuzzy discriminant C-Means clustering model. Journal of Food Process Engineering, 40, 1-7.

Wu, X., Wu, B., Sun, J., Li, M., & Du, H. (2016). Discrimination of Apples Using Near Infrared Spectroscopy and Sorting Discriminant Analysis. International Journal of Food Properties, 19(5), 1016-1028.

Figure captions

Fig. 1 Visual appearance of the apple varieties used in this work.

Fig. 2 VIS-NIR in-line design composed of camera, illumination and linear actuator with reflection probe.

Fig. 3 Linear actuator with reflection probe in (a) resting position, and (b) measuring the fruit, a segmented image of the fruit (Fig. 3a right) and the VIS-NIR spectra obtained (Fig. 3b right).

Fig. 4 Average Log (1/R) of the 50 yellow and red apples at different testing distances.

Fig. 5 Raw VNIR and NIR spectra of five different varieties of apple.

Fig. 6 PCA analysis of apple varieties of the training samples.

Table captions

Table 1 Characteristics of the five apple varieties studied.

Table 2 The contribution and accumulative contribution rates of the first seven PCs.

Table 3 LDA and QDA confusion matrices, both presented as a percentage of correctly

classified samples in the test data set.