Prediction of the astringency level in persimmon using visible and near infrared spectroscopy

Victoria Cortés[1], Alejandro Rodríguez[2], Sergio Cubero[3], José Blasco[3], Cristina Besada[4], Beatriz Rey[2], Pau Talens[1], Alejandra Salvador[4], Nuria Aleixos[2(*)]

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

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

3) Centro de Agroingeniería. Instituto Valenciano de Investigaciones Agrarias (IVIA). Ctra. Moncada-Náquera Km 4.5, 46113, Moncada, Valencia (Spain)

4) Centro de Tecnología Postcosecha. Instituto Valenciano de Investigaciones Agrarias (IVIA). Ctra. Moncada-Náquera Km 4.5, 46113, Moncada, Valencia (Spain)

*E-mail of the corresponding author: naleixos@dig.upv.es

ABSTRACT

Early diagnosis of fruit quality requires a reliable and fast determination. Recently, the industry has an increased interest in non-destructive methods, such as visible and near infrared spectroscopy techniques. The aim of this study was to evaluate the feasibility of visible and near infrared (VIS/NIR) spectroscopy to predict the changes of astringency in intact and the flesh of half cut persimmon fruits in combination with multivariate techniques. The fruits were harvested and exposed to different treatments with 95% CO_2 at 20 ºC during 0 h, 6 h, 12 h, 18 h and 24 h in order to obtain fruit with different levels of astringency. Based on a training dataset of 98 samples, different predictive models were developed from their VIS/NIR spectral data and the subsequent external validation carried out with 42 samples. For modelling, the partial least squares (PLS), support vector machine (SVM) and least squares support vector machine (LS-SVM) were used. The models with best performance were those which included 2-Der, SNV+2-Der, and MSC+2-Der in the pre-processing applied and using the average of the six measuring points in intact fruit. The best model found was the SVM developed with SNV+2-Der pre-processing and the average of the six measuring points of the intact fruit and with all wavelengths (R^2_P = 0.8991 and RPD = 3.1865). Regarding to EWs, with only 16 EWs was achieved a R^2_P of 0.8933 and RPD of 3.0978 for the SVM model using the six measuring points of the intact fruit and 2-Der pre-processing. These results suggest that

VIS/NIR spectroscopy has potential in the non-destructive determination of the astringency in persimmon fruits easily and rapidly without costly and difficult chemical analysis.

Keywords: *Diospyros kaki*, fruit internal quality, soluble tannins, near infrared spectroscopy, chemometrics

1. INTRODUCTION

The persimmon (*Diospyros kaki* L.) is a fruit native to China, but now cultivated in warm regions worldwide[1]. Climate characteristics of the production area are important factors influencing the quality and features of the fruits. The main land areas in which this type of fruit is grown in Spain are La Ribera del Xúquer with European Protected Designation of Origin (PDO) register label, Alicante, Andalucía, Castellón, Extremadura and Valencia[2]. There are several different persimmon cultivars farmed in Spain; such as 'Rojo Brillante' astringent type. Persimmon develops an astringent taste due to the presence of soluble tannins. Tannins are polyphenol compounds with a high molecular weight. Their large hydroxyl phenolic groups cause astringency. As the fruit ripens, the soluble tannins gradually change to insoluble tannins; this makes the fruit tastes less astringent[3]. However, there are several postharvest treatments to achieve the fast elimination of the astringency of the fruits without the pulp firmness is affected[4]. Among them, the technique most commercially used is based on exposing fruits to high CO_2 concentration (95% - 98%). This method promotes anaerobic respiration in the fruit, giving rise to an accumulation of acetaldehyde which reacts with the soluble tannins. Tannins become insoluble with the treatment and astringency removed[5]. If the treatment is too short, it may result on fruit with residual astringency[6], and if it is excessively extended may lead to fruit quality losses[7] and therefore non-destructive techniques to ensure the success of the treatments are searched.

Spectroscopy techniques have been widely used for qualitative and quantitative determination of different compounds in fruit samples. One of the spectroscopy techniques used in food chemistry is near infrared (NIR) spectroscopy[8-9-10] because it is non-destructive, inexpensive, rapid and reliable. This technique has been utilized for the quantitative determination of total soluble solids, firmness, acidity, dry matter, chemical substances such as glucose, sucrose, citric acid, malic acid, starch or cellulose[11-12-13],

and even to determine a maturity index[14], internal quality index[15] or different appropriate indices for quality analysis[16].

Versatile combination of chemometrics and VIS/NIR spectroscopy has been applied in food industry, agriculture and horticulture to obtain information from spectra. Support vector machines (SVM) are learning algorithms used for classification and regression tasks widely used in the analysis of spectroscopic data[17-18]. Chauchard, et al., (2004)[19] compared classical linear regression techniques and Least Square Support Vector Machine (LS-SVM) regression for the prediction of total acidity in fresh grapes using NIR spectroscopy. LS-SVM increased prediction used in combination with standard normal variate (SNV) pre-processing and partial least square regression (PLSR) latent variables. Nicolaï et al., (2007)[20] predicted the sugar content using PLS. Covariance, Gaussian and cubic polynomial kernel functions were considered obtaining similar results for all methods about R^2=0.87 and Q^2=0.84, concluding that kernel PLS does not offer advantages compared to ordinary PLS. The identification of the spectral variables that provide usefulness can lead to better results in classification and simplifies the chemical interpretation of the results. Calvini et al., (2015)[21] tested sparse principal component analysis (PCA) coupled with k-Nearest-Neighbours (k-NN) and sparse PLS-DA, to discriminate Arabica and Robusta coffee, and compared the results with the classical approaches based on PCA+kNN and PLS-DA. Folch-Fortuny et al., (2016)[22] used N-way-PLS-DA to early detect invisible decay lesions in citrus fruit achieving a prediction above 90%.

Some researchers have used NIR spectroscopy for non-destructive quality analysis of persimmons. Mowat and Poole (1997)[23] concluded that NIR spectroscopy has been a beneficial technique in determination of the quality of persimmon fruit. Ito et al. (1997)[24] and Noypitak et al. (2014)[3] investigated astringency in the "Nisimura-wase" and in 'Xichu' persimmons, respectively. The most common mode is the diffuse reflectance, which acquires the reflected light in the vicinity of the illuminating point and is preferable for the measurement of intact fruit[25-26]. The aim of this study was to evaluate the usefulness of chemometrics and VIS/NIR spectroscopy to create a non-destructive tool to estimate/predict the level of astringency in persimmons cv. 'Rojo Brillante'.

2. Plant material and experimental design

Persimmon cv. 'Rojo Brillante' fruits were harvested in L´Alcudia (Valencia, Spain) at two different commercial maturity stages (M1 and M2) corresponding to end of November and mid of December. [27]The maturity index used to select the fruits was a visual observation of the external colour of the fruit. After each harvest, 70 fruits with not skin damage and homogenous colour were selected (making a total of 140 fruits). The colour index (CI=100*a*/*Lb*, Hunter parameters) at harvest was 18.20 ± 3.32 and 21.6 ± 4.05 for the stages M1 and M2, respectively. Fruit firmness was also determined by means of a universal testing machine (4301, Instron Engineering Corp., MA, USA) equipped with an 8 mm puncture probe. The crosshead speed during firmness testing was 10 mm/min. During the test, the force increased smoothly until it decreased drastically when the flesh was broken, and then the maximum peak force was registered. The results were expressed as the load (in N) required for breaking the flesh of the fruit on both sides after peel removal. In the samples used for these experiments, fruit firmness decreased together with the stage of maturity at harvest, being the mean values of firmness 30.8 N ± 3.5 and 24.4 N ± 4.9 for the stages M1 and M2, respectively.

In order to obtain different levels of astringency, fruit at each maturity stage was divided into five homogeneous lots. Then fruit was exposed to CO_2 treatments in closed containers (95% CO_2 at 20 ºC and 90% RH) for: 0 h, 6 h, 12 h, 18 h and 24 h. Spectroscopic measurements of the intact fruits and the internal flesh of half cut fruits were acquired in the 8 h after the CO_2 treatment. The Figure 1 shows the location of the selected points for the measurements.

Figure 1. Selected points for the spectroscopic measurements in the: a) intact fruit; and b) the flesh of half cut fruit, respectively

The level of astringency of each individual fruit was determined in the following way. The flesh sample of each fruit was frozen at −20 ºC and later the content of soluble tannins was analysed using the Folin-Denis method[28]. The results were expressed as relative soluble tannins in fresh weight. Before this process, each fruit was additionally cut in half and pressed against 10x10 cm filter paper soaked on a 5% $FeCl_3$ solution, obtaining a print whose quantity and intensity gave information about the soluble tannins content and its distribution[29]. This tannin print method is an alternative

technique to the Folin-Denis method used in industry on random fruits to determine the level of astringency in fruit batches.

3. Visible and near-infrared spectra collection

The VIS/NIR spectra were collected alternately in reflectance mode using a multichannel spectrometer platform (AVS-DESKTOP-USB2, Avantes BV, The Netherlands) equipped with two detectors (Fig. 2). The first detector (AvaSpec-ULS2048 StarLine, Avantes BV, The Netherlands) included a 50 mm entrance slit and a 600 lines/mm diffraction grating covering the VIS/NIR range from 600 nm to 1100 nm with a spectral FWHM (full width at half maximum) resolution of 1.15 nm. The spectral sampling interval was 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), a 100 mm entrance slit and a 200 lines/mm diffraction grating covering the NIR range of 900 nm to 1800 nm and a spectral FWHM resolution of 12 nm. The spectral sampling interval was 3.535 nm. A stabilised 10 W tungsten halogen light source (AvaLight-HAL-S, Avantes BV, The Netherlands) was used. By using a holder the sample was properly positioned over the probe and the reflectance probe delivered the light to the sample and collected the reflectance from the sample, which was carried by the fibre cable to the spectrometer in use. The probe tip was designed to provide reflectance measurements at an angle of 45º so as to minimise specular reflectance from the surface of the fruit.

The calibration was performed using a 99% reflective white reference tile (WS-2, Avantes BV, The Netherlands) so that the maximum reflectance value over the range of wavelengths was around 90% of saturation. Prior to spectral measurements, the temperature of the persimmons was stabilised at 24 ºC. Measurements were taken at the six different points over the surface of the intact persimmon and the flesh of half cut fruit (Fig. 1) and average values of the spectra for both types of measurements were used for the analysis. A personal computer equipped with commercial software (AvaSoft version 7.2, Avantes, Inc.) was used to control both detectors and to acquire the spectra. The signals were pre-processed using AvaSoft software. The integration time was set to 90 ms for the detector sensitive in the VIS/NIR region and to 700 ms for the detector sensitive in the NIR region. For both detectors, each spectrum was obtained as the average of five scans to reduce the thermal noise of the detector[20]. The average

reflectance measurements of each sample (S) were then converted into relative reflectance values (R) with respect to the white reference using dark reflectance values (D) and the reflectance values of the white reference (W), as shown in (1):

$$R = \frac{S-D}{W-D} \quad (1)$$

The dark spectrum was obtained by turning off the light source and completely covering the tip of the reflectance probe.

Figure 2. A labelled picture of the VIS/NIR equipment

4. Statistical analysis

The VIS/NIR data and the tannins values were organised into matrices, in which the rows represent the samples (total number of samples is 140 persimmons) and the columns represent the variables (X-variables and Y-variable). The X-variables, or predictors, were the concatenation of the VIS spectra (600-1100 nm) and the NIR spectra (900 – 1800 nm) for each persimmon (globally, 2248 wavelengths). The Y-variable, or response, in the last column of the matrices, represented the measured tannin value associated to each sample.

Fourteen matrices were generated corresponding to different combinations of the measurements from the intact fruit and the flesh of the half cut fruit. The first matrix corresponded to the mean reflectance values of the measurements in the six points (1 to 6) from the intact fruit depicted in Fig. 1. The second matrix contained mean values of the measurements in four points (2-5-3-4), which correspond to the lowest part of the persimmon, for the intact fruit. The third to seventh matrices contained mean values of measurements from other combinations of points (1-6-2-5, 1-6-3-4, 1-6, 2-5 and 3-4). The remaining seven matrices corresponded to the mean values of the measurements from the same combinations of points, but from the flesh of the half cut fruit.

4.1. Spectral Pre-processing

In order to remove the influence of undesired effects, like high-frequency noise, baseline shifts, light scattering, random noise and any other external effects, either due to instrumental or ambient factors, seven different spectral pre-processing methods and their combinations were separately applied prior to the development of the prediction models. These methods included SNV, multiplicative scatter correction (MSC),

detrending, Savitzky-Golay smoothing (SG), first (1-Der) and second (2-Der) derivatives, and direct orthogonal signal correction (DOSC). All spectral pre-processing methods and the development of the prediction models were carried out using MATLAB R2015b (The Mathworks Inc., Natick, MA, USA) and its toolboxes as custom MATLAB scripts.

[30]SNV is used to eliminate the multiplicative noise due to particle size influence or scatter interference. [31]SNV subtracts the mean from an individual spectrum and divides it by its standard deviation. [32-33]Similarly, the MSC is used to compensate the effect of non-uniform scattering induced by diverse particle sizes and other physical effects in the spectral. It linearises each spectrum to a mean spectrum (derived from the calibration set) and fits it using the least squares method.

On the other hand, detrending is a polynomial baseline correction method that is used to eliminate the baseline shifts in the spectrum through the subtraction of the linear trend from the original spectrum[34-31]. Smoothing is an effective way for reducing the high-frequency noise. The average of several points in a window is calculated in order to obtain an optimal estimation. Several smoothing methods exist in the literature, but one of the most commonly applied is the Savitzky-Golay (SG)[35] smoothing. This method has the advantage of preserving the characteristics of the signal such as the maximum and minimum relative values, as well as the width of the peaks, which usually disappear with other smoothing methods. In the present work, the SG smoothing was calculated with two-degree polynomials and a window size of seven points.

[36]The first (1-Der) and second (2-Der) derivatives are well-accepted pre-processing methods to eliminate the shifting, the scattering and the background noise as well as to distinguish overlapped peaks and to improve the spectral resolution. [37]They were calculated using the SG algorithm with three point smoothing filters and two-degree polynomial.

Finally, [38]DOSC methods are used to remove information that has little correlation (orthogonal) with the response matrix. DOSC obtains components that are orthogonal to the response matrix and deletes those that are considered irrelevant, thus improving the prediction ability.

4.2. Modelling by different calibration methods

Prior to applying the pre-processing, each set was randomly divided into two groups, a calibration set (70% of the samples, 98 persimmons) and a prediction set (30% of the

samples, 42 persimmons) with the purpose of evaluating the reliability of three different regression techniques used to predict the tannins content. The three regression techniques used in this work are: PLSR, SVM, and the LS-SVM regression.

[39]PLS is a widely applied multivariate calibration method for assessing linear relationship between inputs (spectral data or X-variables) and the response variable (tannins content or Y-variable) in studies of spectroscopic analysis. The procedure is based on the use of latent variables (LVs), instead of real variables (spectral data), depending on the covariance between the predictors or X-variables and the response or Y-variable, leading to a parsimonious model with reliable predictive power[40]. In the development of the PLS regression model, 10-fold cross-validation was used to validate its quality, to prevent over-fitting, to choose the number of components that explain most of the variance in the tannins content in calibration and to minimise the expected error when predicting the response from future observations.

[41]SVM method is a popular machine learning tool for regression. It is based on the Vapnik-Chervonenkis (VC) dimension and the structural risk minimisation principle[42]. It is considered as a nonparametric technique because SVM models are based on a non-linear kernel function. In short, SVM maps the calibration dataset to a high-dimensional feature space by a non-lineal mapping, to subsequently perform a linear regression. The SVM regression has the advantage of being very efficient and robust during the training model. In this study, the statistical and machine learning toolbox of Matlab was used to train a SVM model with the spectral and tannins information using a linear kernel and a 10-fold cross-validation.

[43-44]LS-SVM is a learning algorithm which improves the generalisation ability of the machine learning procedure based on the principle of structural risk minimisation. It handles both linear and nonlinear multivariate problems with less computational cost and with a small sample database. This is achieved using linear equations instead of quadratic problems to reduce the complexity of the optimisation process[45]. The LS-SVM has the advantage of limited over-fitting, high predictive reliability and strong generalisation ability. The LS-SVMlab v1.8 toolbox (Suykens, Leuven, Belgium) was applied to develop the calibration models. During the calibration, the linear kernel and a 10-fold cross-validation were applied to avoid over-fitting problems. The linear kernel included a regulation parameter that determined the trade-off between minimising the training error and minimising the model complexity. Large γ implies little regularisation, and thus a more nonlinear model[46].

4.3. Variable selection

As the 2248 variables employed as inputs (wavelengths) in the models might contain much collinearity and redundancy, some wavelengths were selected as effective wavelengths (EW) for each model. This was made with the purpose of reducing the high dimensionality of the spectral data and the computational cost, thus achieving an optimal model.

The algorithm that was applied to select the EWs was a successive projection algorithm (SPA). [47-48-49]SPA is a variable selection algorithm applied to solve collinearity problems and to select the EWs with least redundancies by means of a simple projection procedure in a vector space. The best subsets of wavelengths that fit the minimum collinearity were selected. For each calibration model, SPA was applied and the obtained EWs were used again as the inputs of the PLS, SVM and LS-SVM models.

4.4. Model evaluation

The accuracy and the predictive capability of the three different models were assessed by means of the coefficient of determination (R^2) and the correlation coefficient (ρ), for both calibration and prediction, the root mean square error of calibration (RMSEC), the root mean square error of prediction (RMSEP) and the ratio of performance to deviation (RPD). Generally, a good model should have high R^2 and a correlation coefficient with low RMSEC and RMSEP value. In addition, an acceptable model should have a RPD value of more than 2.5, being very good above 3.0[50-51-52-15]. These parameters can be defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(\hat{y}_i - \bar{y}_i)^2} \quad (2)$$

$$RMSEC\ or\ RMSEP = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \quad (3)$$

$$RPD = \frac{SD(y)}{RMSEP} \quad (4)$$

Where:

$\hat{y}$ is the predicted value of the i^{th} persimmon.

y_i is the measured value of the i^{th} persimmon.

N: is the number of observations in the calibration set or in the prediction set.

SD: is the standard deviation of the measured values.

5. RESULTS AND DISCUSSION

The total number of persimmon samples was 140 with a mean tannins content of 0.2497 (STD=0.2208). The statistical values of tannins content of persimmons in calibration and prediction sets are shown in Table 1.

Table 1. Statistical values of tannins content of persimmons

Before applying the models, the raw reflectance spectra (Fig. 3) of the samples were treated by the previously described pre-processing methods.

Figure 3. Raw reflectance spectra (%) of the persimmons calibration set for: (a) the VIS/NIR region; and (b) the NIR region

Thus, the PLS, SVM and LS-SVM models were developed with raw and pre-processed spectra. The samples in the prediction set were used to assess the prediction performance of developed models, and according to the indicated evaluation standards, such as R^2, ρ, RMSE and RPD, the optimal pre-processing methods could be determined.

The prediction results by the models are shown in Table 2 for the average of the six measuring points with the original spectra for the intact fruit set, and in Table 3 for the half cut fruit set, respectively.

Table 2. Prediction results of tannins content using the average of the six measuring points with all wavelengths by PLS, SVM and LS-SVM models for the intact fruit set

Table 3. Prediction results of tannins content using the average of the six measuring points with all wavelengths by PLS, SVM and LS-SVM models for the half cut fruit set

From the Tables 2-3 we can see that, as average, the models with best performance are those which included 2-Der in the pre-processing applied (2-Der, SNV+2-Der, MSC+2-Der).

Tables 4-5 show the results for the three selected methods and the aforementioned pre-processing combinations after applying the SPA methods for wavelength selection. The selected EWs by SPA are shown in Table 6. According to the SPA, the EWs were shown in order of importance.

Table 4. Prediction results of tannins content using the average of the six measuring points with EWs (by SPA) by PLS, SVM and LS-SVM models for the intact fruit set

Table 5. Prediction results of tannins content using the average of the six measuring points with EWs (by SPA) by PLS, SVM and LS-SVM models for the half cut fruit set

Table 6. Selected effective wavelengths (EWs) from the average of the six measuring points by SPA for the intact fruit set and the half cut fruit set

This analysis was performed for the different combinations of the six measuring points tested obtaining the results shown in Table 7, which shows the best results for each combination of points and each model. Tables 8-9 show the complete results for the combination of the 2-5-3-4 measuring points (average of the equator and bottom measurements) for the intact fruit set and for the half cut fruit set, respectively, which correspond to the best obtained results. The highest RPD with the different models with this combination of measuring points is always equal or better than the highest RPD obtained with the other combinations of points. This is reasonable, since from the tannin prints observed in Fig. 4, obtained using the technique based on $FeCl_3$, the higher differences are in the equator-bottom part, being the upper part more similar in fruits with different CO_2 treatments (Fig. 4b-e).

Figure 4. Tannin prints representing the evolution of the astringency distribution and intensity for persimmons under different CO_2 treatments: a) non-treated; and b-e) treated with CO_2 for 6 h, 12 h, 18 h and 24 h, respectively

Table 7. Prediction results of tannins content using different combinations of measuring points and pre-processing with all wavelengths by PLS, SVM and LS-SVM models*

* Only the best prediction results for each model are shown, indicating the associated pre-processing

Table 8. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with all wavelengths by PLS, SVM and LS-SVM models for the intact fruit set

Table 9. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with all wavelengths by PLS, SVM and LS-SVM models for the half cut fruit set

As in the previous case, the observed models with best performance were chosen for wavelengths reduction. Again, from the Tables 8-9 it can be observed that the models with best performance are those which included 2-Der in the pre-processing applied (2-Der, SNV+2-Der, and MSC+2-Der). However, in the case of the half cut case also the 1-Der pre-processing obtained a good prediction and was therefore included in the study.

Tables 10-11 show the results for the selected methods and the aforementioned pre-processing combinations after applying the SPA method for wavelength selection. The selected EWs are shown in Table 12 in order of importance according to the SPA.

Table 10. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with EWs (selected by SPA) by PLS, SVM and LS-SVM models for the intact fruit set

Table 11. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with EWs (selected by SPA) by PLS, SVM and LS-SVM models for the half cut fruit set

Table 12. Selected effective wavelengths (EWs) from the average of the four measuring points (2-5-3-4) by SPA for the intact fruit set and the half cut fruit set

In this work, different models were obtained using different spectra pre-processing, statistical methods and number of effective wavelengths. For the intact fruit and the half cut fruit, the PLSR method obtained worse predictions than SVM and LS-SVM methods. In the three aforementioned methods the average of the six and four measuring points were used, being the best pre-processing found the 2-Der, SNV+2-Der and MSC+2-Der (those where the second derivative has been used), for both, the models created with the full spectral range and for the models selecting EWs.

In addition, for the average of the six and four measuring points of the intact fruit, less EWs were selected. The best model found was the SVM with SNV+2-Der pre-processing using the average of the six measuring points of the intact fruit and all wavelengths (R^2_P=0.8991; RPD=3.1865). Regarding to EWs, the best model with the minimum number of wavelengths (EWs=16) was the SVM model using the six measuring points of the intact fruit and 2-Der pre-processing (R^2_P=0.8933; RPD=3.0978).

6. CONCLUSIONS

This study states the VIS/NIR spectroscopy as a suitable non-destructive method to determine the astringency in persimmon fruit easily and rapidly without costly and tedious chemical analysis or subjective evaluation by the tannin print assessment. Spectra of selected points in intact and half cut persimmons were acquired in the visible and near infrared regions. A total of seven signal pre-processing methods including SNV, detrending, SG, 1-Der, 2-Der, MSC, DOSC, and combinations of them have been used on the measurements of the single and combination of the selected points. The combinations considered were SNV+2-Der and MSC+2-Der, since they were observed in our preliminary studies as those with the best performance from all the combinations evaluated (notice the 2-Der is involved in these two combinations). Astringency in persimmon fruits was predicted using three regression techniques like PLSR, SVM, and LS-SVM.

In addition, effective wavelengths were obtained using the SPA. Depending on the method, the effective wavelengths vary from 8 to 18 when the single points are used

and from 13 to 39 when the combination 2-5-3-4 (that was the best combination) was used. In general, EWs belonged to the NIR region, ranging from 880 to 1100 nm with a few exceptions.

The best astringency prediction model was obtained with SVM method, after pre-processing SNV+2-Der of the full spectra, using the average of the six selected points of the intact fruit reaching R^2_P=0.8991 and RPD=3.1865. Moreover, the best prediction results by selecting 16 effective wavelengths was for the SVM model using the six measuring points of the intact fruit and 2-Der pre-processing (R^2_P=0.8933; RPD=3.0978).

Thus, these results point the VIS/NIR spectroscopy as a feasible technique to determine the astringency in persimmon fruits easily and rapidly non-destructively.

ACKNOWLEDGEMENTS

This work has been partially funded by the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria de España (INIA) through research projects RTA2012-00062-C04-01, RTA2012-00062-C04-03 and RTA2013-00043-C02 with the support of European FEDER funds and by the Conselleria d' Educació, Investigació, Cultura i Esport, Generalitat Valenciana, through the project AICO/2015/122. V. Cortés thanks the Spanish MEC for the FPU grant (FPU13/04202).

REFERENCES

1. Ashtiani, S.M., Salarikia, A., Golzarian, M.R. & Emadi, B. Non-Destructive Estimation of Mechanical and Chemical Properties of Persimmons by Ultrasonic Spectroscopy. *International Journal of Food Properties* **2016**, 19:7, 1522-1534.

2. Khanmohammadi, M., Karami, F., Mir-Marqués, A., Bagheri Garmarudi, A., Garrigues, S. & de la Guardia, M. Classification of persimmon fruit origin by near infrared spectrometry and least squares-support vector machines. *Journal of Food Engineering* **2014**, 142, 17-22.

3. Noypitak, S., Terdwongworakul, A., Krisanapook, K. & Kasemsumran, S. Evaluation of astringency and tannin content in 'Xichu' persimmons using near infrared spectroscopy. *International Journal of Food Properties* **2014**, 18:5, 1014-1028.

4. Khademi, O., Mostofi, Y., Zamani, Z. & Fatahi, R. The effect of deastringency treatments on increasing the marketability of persimmon fruit. *Acta Horticulturae* **2010**, 877, 687–691.

463 5. Matsuo, T., Ito, S. & Ben-Arie, R. A model experiment for elucidating the
464 mechanism of astringency removal in persimmon fruit using respiration inhibitors.
465 *Journal of the Japanese Society for Horticultural Science* **1991**, 60, 437–442.

466 6. Besada, C., Salvador, A., Arnal, L. & Martínez-Jávega, J.M. Optimization of the
467 duration of deastringency treatment depending on persimmon maturity. *Acta*
468 *Horticulturae* **2010**, 858, 69–74.

469 7. Novillo, P., Salvador, A., Llorca, E., Hernando, I. & Besada, C. Effect of CO2
470 deastringency treatment on flesh disorders induced by mechanical damage in
471 persimmon. Biochemical and microstructural studies. *Food Chemistry* **2014**, 145, 453 –
472 463.

473 8. Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I. &
474 Lammertyn, J. Non-destructive measurement of fruit and vegetable quality by means of
475 NIR spectroscopy: A review. *Postharvest Biology and Technology* **2007**, 46 (2), 99–
476 118.

477 9. Vitale, R., Bevilacqua, M., Bucci, R., Magrì, A.D., Magrì, A.L., Marini, F. A
478 rapid and non-invasive method for authenticating the origin of pistachio samples by
479 NIR spectroscopy and chemometrics. *Chemometrics and Intelligent Laboratory Systems*
480 **2013**, 121, 90-99.

481 10. Lopez, A., Arazuri, S., Garcia, I., Mangado, J., Jaren, C. A review of the
482 application of near-infrared spectroscopy for the analysis of potatoes. *Journal of*
483 *agricultural and food chemistry* **2013**, 61, 5413–5424.

484 11. Schmilovitch, Z., Mizrach, A., Hoffman, A., Egozi, H. & Fuchs, Y.
485 Determination of mango physiological indices by near-infrared spectrometry.
486 *Postharvest Biology and Technology* **2000**, 19(3), 245-252.

487 12. Nagle, M., Mahayothee, B., Rungpichayapichet, P., Janjai, S. & Müller, J. Effect
488 of irrigation on near-infrared (NIR) based prediction of mango maturity. *Scientia*
489 *Horticulturae* **2010**, 125(4), 771-774.

490 13. Theanjumpol, P., Self, G., Rittiron, R., Pankasemsu, T., & Sardsud, V.
491 Selecting Variables for Near Infrared Spectroscopy (NIRS) Evaluation of Mango Fruit
492 Quality. *Journal of Agricultural Science* **2013**, 5(7).

493 14. Jha, S.N., Jaiswal P, Narsaiah K, Sharma R, Bhardwaj R, Gupta M, Kumar R.
494 Authentication of mango varieties using near infrared spectroscopy. *Agricultural*
495 *Research* **2013**, 2(3), 229 – 235.

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

16. Attila, N. & János, T. Sweet cherry fruit analysis with reflectance measurements. *Journal Analele Universităţii din Oradea, Fascicula: Protecţia Mediului* **2011**, 17, 263-270.

17. Devos, O., Ruckebusch, C., Durand, A., Duponchel, L., Huvenne, J.P. Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation. *Chemometrics and Intelligent Laboratory Systems* **2009**, 96, 27-33.

18. Fernández Pierna, J.A., Vermeulen, P., Amand, O., Tossens, A., Dardenne, P., Baeten, V. NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. *Chemometrics and Intelligent Laboratory Systems* **2012**, 117, 233-239.

19. Chauchard, F., Cogdill, R., Roussel, S., Roger, J.M., Bellon-Maurel, V. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems* **2004**, 71, 141-150.

20. Nicolaï, B.M., Theron, K.I., Lammertyn, J. Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple. *Chemometrics and Intelligent Laboratory Systems* **2007**, 86, 243–252.

21. Calvini, R., Ulrici, A., Amigo, J.M. Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems* **2015**, 146, 503–511.

22. Folch-Fortuny, A., Prats-Montalbán, J.M., Cubero, S., Blasco, J., Ferrer, A. NIR hyperspectral imaging and N-way PLS-DA models for detection of decay lesions in citrus fruits. *Chemometrics and Intelligent Laboratory Systems* **2016**, 156, 241-248.

23. Mowat, A.D. & Poole, P.R. Non-destructive discrimination of persimmon fruit quality using visible-near infrared reflectance spectrophotometry. *Acta Horticulturae* **1997**, 436, 159–164.

24. Ito, S., Ootake, Y. & Kito, I. (1997). Classification of astringency in pollination variant non-astringent persimmon fruits cv. “Nisimura-wase” by near infrared spectroscopy. Research Bulletin of the Aichi-ken Agricultural Research Center, 29, 213–218.

25. Shao, Y., He, Y., Bao, Y. & Mao, J. Near-infrared spectroscopy for classification of oranges and prediction of the sugar content. *International Journal of Food Properties* **2009**, 12 (3), 644–658.

26. He, Y., Li, X. & Shao, Y. Fast discrimination of apple varieties using vis/nir spectroscopy. *International Journal of Food Properties* **2007**, 10 (1), 9–18.

27. Salvador, A., Arnal, L., Besada, C., Larrea, V., Quiles, A. & Pérez-Munuera, I. Physiological and structural changes during ripening and deastringency treatment of persimmon cv. 'Rojo Brillante'. *Postharvest Biology and Technology* **2007**, 46, 181–188.

28. Taira, S. (1995). Astringency in persimmon. In: Linskens, H.F., Jackson, J.F. Fruit Analysis. Springer, Hannover, Germany, pp. 97–110.

29. Matsuo, T. & Ito, S. A model experiment for de-astringency of persimmon fruit with high carbon dioxide: in vitro gelation of kaki-tannin by reacting with acetaldehyde. *Journal of Agricultural Food Chemistry* **1982**, 46, 683–689.

30. Rinnan, Å., van den Berg, F., & Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry* **2009**, 28(10), 1201-1222.

31. Feng, Y. Z., & Sun, D. W. Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of Pseudomonas loads in chicken fillets. *Talanta* **2013**, 109, 74-83.

32. Fearn T, Riccioli C, Garrido-Varo, A, Guerrero-Ginel, J.E. On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems* **2009**, 96, 22–26.

33. Vidal, M., & Amigo, J. M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems* **2012**, 117, 138-148.

34. Zhang, L., Sun, B., Xie, P., Li, H., Su, H., Sha, K., Huang, C., Lei, Y., Liu, X., Wang, H. Using near infrared spectroscopy to predict the physical traits of Bos grunniens meat. *LWT - Food Science and Technology* **2015**, 64 (2), 602-608.

35. Savitzky, A., & Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* **1964**, 36(8), 1627-1639.

36. Sinija, V. R., & Mishra, H. N. FTNIR spectroscopic method for determination of moisture content in green tea granules. *Food and Bioprocess Technology* **2011**, 4(1), 136-141.

37. Liu, Y., Sun, X., Zhou, J., Zhang, H., & Yang, C. Linear and nonlinear multivariate regressions for determination sugar content of intact Gannan navel orange by Vis–NIR diffuse reflectance spectroscopy. *Mathematical and Computer Modelling* **2010**, 51(11), 1438-1443.

38. Zhu, D., Ji, B., Meng, C., Shi, B., Tu, Z., & Qing, Z. The application of direct orthogonal signal correction for linear and non-linear multivariate calibration. *Chemometrics and Intelligent Laboratory Systems* **2008**, 90(2), 108-115.

570 39. Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial.
571 *Analytica chimica acta* **1986**, 185, 1-17.

572 40. Lorber, A., Wangen, L., Kowalski, B. A theoretical foundation for the PLS
573 algorithm. *Journal of Chemometrics* **1987**, 1, 19–31

574 41. Vapnik, V. (2013). The nature of statistical learning theory. Springer Science &
575 Business Media.

576 42. Gunn, S. R. (1998). Support vector machines for classification and regression.
577 ISIS technical report, 14.

578 43. Liu, F., He, Y., & Wang, L. Comparison of calibrations for the determination of
579 soluble solids content and pH of rice vinegars using visible and short-wave near infrared
580 spectroscopy. *Analytica chimica acta* **2008**, 610(2), 196-204.

581 44. Suykens, J. A., & Vandewalle, J. Least squares support vector machine
582 classifiers. *Neural processing letters* **1999**, 9(3), 293-300.

583 45. Liu, F., He, Y., & Sun, G. Determination of protein content of Auricularia
584 auricula using near infrared spectroscopy combined with linear and nonlinear
585 calibrations. *Journal of agricultural and food chemistry* **2009**, 57(11), 4520-4527.

586 46. Sun, T., Lin, H., Xu, H., & Ying, Y. Effect of fruit moving speed on predicting
587 soluble solids content of 'Cuiguan'pears (Pomaceae pyrifolia Nakai cv. Cuiguan) using
588 PLS and LS-SVM regression. *Postharvest biology and technology* **2009**, 51(1), 86-90.

589 47. Araújo, M. C. U., Saldanha, T. C. B., Galvão, R. K. H., Yoneyama, T., Chame,
590 H. C., & Visani, V. The successive projections algorithm for variable selection in
591 spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory*
592 *Systems* **2001**, 57(2), 65-73.

593 48. Galvao, R. K. H., Araujo, M. C. U., Fragoso, W. D., Silva, E. C., Jose, G. E.,
594 Soares, S. F. C., & Paiva, H. M. A variable elimination method to improve the
595 parsimony of MLR models using the successive projections algorithm. *Chemometrics*
596 *and intelligent laboratory systems* **2008**, 92(1), 83-91.

597 49. Zhang, S., Zhang, H., Zhao, Y., Guo, W., & Zhao, H. A simple identification
598 model for subtle bruises on the fresh jujube based on NIR spectroscopy. *Mathematical*
599 *and Computer Modelling* **2013**, 58(3), 545-550.

600 50. Williams, P., Sobering, D. Comparison of commercial near infrared
601 transmittance and reflectance instruments for analysis of whole grains and seeds.
602 *Journal of Near Infrared Spectroscopy* **1993**, 1, 25.

603 51. Viscarra Rossel, R. A., Taylor, H. J., & McBratney, A. B. Multivariate
604 calibration of hyperspectral γ□ray energy spectra for proximal soil sensing. *European*
605 *Journal of Soil Science* **2007**, 58(1), 343-353.

606 52. Kamruzzaman, M., Makino, Y., & Oshita, S. Rapid and non-destructive
607 detection of chicken adulteration in minced beef using visible near-infrared
608 hyperspectral imaging and machine learning. *Journal of Food Engineering* **2016**, 170,
609 8-15.

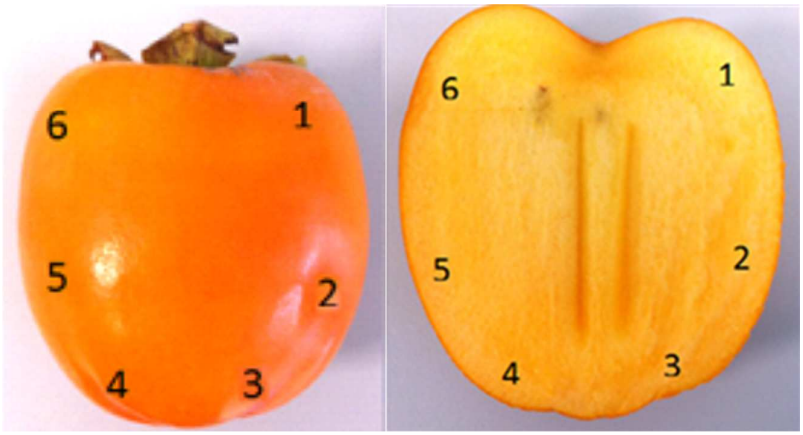

Figure 1. Selected points for the spectroscopic measurements in the: a) intact fruit; and b) the flesh of half cut fruit, respectively

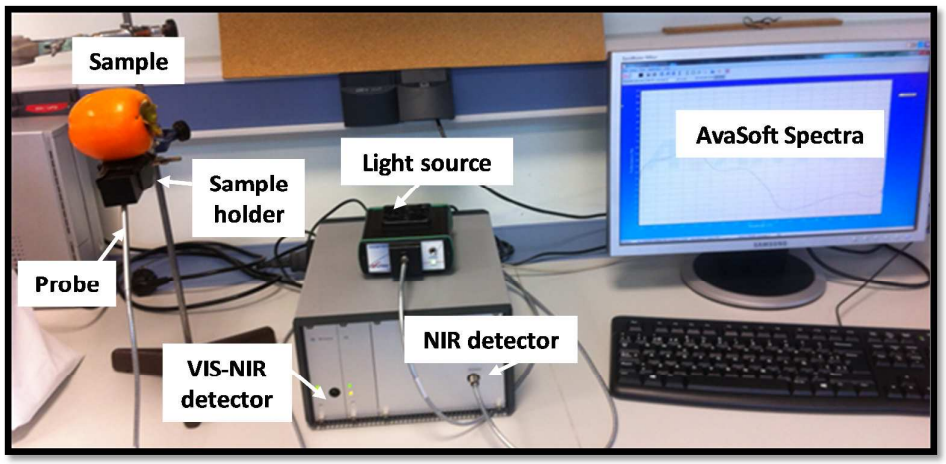

Figure 2. A labelled picture of the VIS/NIR equipment

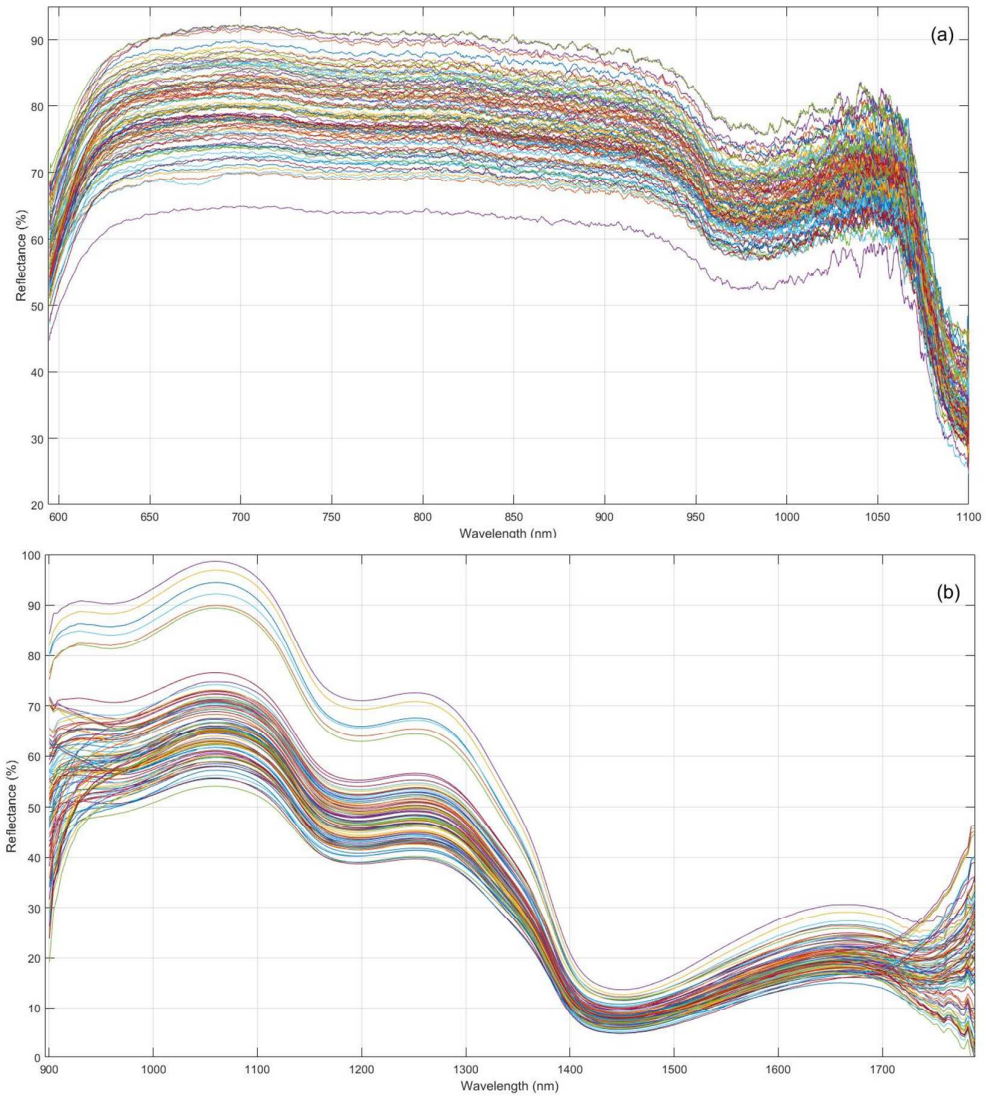

Figure 3. Raw reflectance spectra (%) of the persimmons calibration set for: (a) the VIS/NIR region; and (b) the NIR region

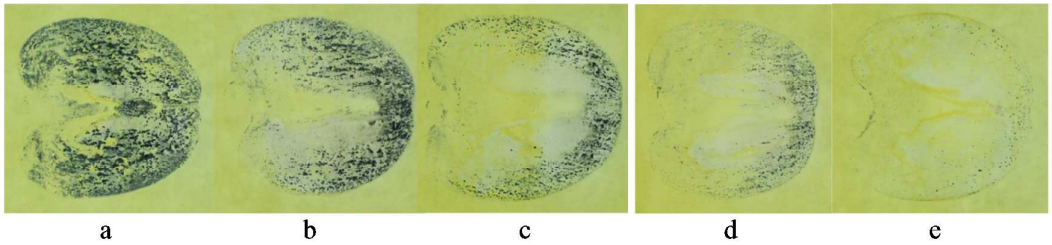

Figure 4. Tannin prints representing the evolution of the astringency distribution and intensity for persimmons under different CO_2 treatments: a) non-treated; and b-e) treated with CO_2 for 6 h, 12 h, 18 h and 24 h, respectively

Table 1. Statistical values of tannins content of persimmons

DATA SET	Sample Nº	Min	Max	Mean	STD
Calibration	98	0.023	0.735	0.243	0.210
Prediction	42	0.023	0.752	0.266	0.245

Table 2. Prediction results of tannins content using the average of the six measuring points with all wavelengths by PLS, SVM and LS-SVM models for the intact fruit set

MODEL	PRE-PROCESSING	*LV*, γ	Calibration			Prediction			
			R^2	ρ	RMSE	R^2	ρ	RMSE	RPD
PLSR	**RAW**	20	0,992	0,996	0,018	0,762	0,887	0,118	2,08
	SNV	19	0,992	0,996	0,019	0,817	0,913	0,104	2,36
	Detrend	18	0,990	0,995	0,021	0,792	0,902	0,111	2,22
	SG	21	0,991	0,995	0,020	0,771	0,892	0,116	2,12
	1-Der	12	0,992	0,996	0,018	0,796	0,906	0,120	2,24
	2-Der	14	0,993	0,996	0,018	0,795	0,894	0,120	2,23
	MSC	18	0,991	0,996	0,020	0,830	0,919	0,100	2,45
	DOSC	1	1,000	1,000	0,000	0,745	0,879	0,123	2,00
	SNV + 2-Der	14	0,992	0,996	0,018	0,820	0,908	0,103	2,39
	MSC+2- Der	14	0,993	0,996	0,018	0,823	0,909	0,102	2,41
SVM	**RAW**		0,812	0,916	0,091	0,742	0,899	0,123	1,99
	SNV		0,939	0,969	0,052	0,853	0,931	0,093	2,64
	Detrend		0,932	0,966	0,055	0,842	0,927	0,096	2,55
	SG		0,815	0,918	0,090	0,715	0,894	0,129	1,90
	1-Der		0,971	0,987	0,035	0,879	0,945	0,084	2,91
	2-Der		0,972	0,988	0,035	0,891	0,946	0,080	3,06
	MSC		0,945	0,972	0,049	0,867	0,940	0,088	2,77
	DOSC		0,993	1,000	0,017	0,763	0,880	0,118	2,08
	SNV + 2-Der		0,973	0,989	0,035	0,899	0,955	0,077	3,19
	MSC+2- Der		0,973	0,988	0,034	0,8944	0,951	0,079	3,10
LS-SVM	**RAW**	0,210	0,944	0,974	0,050	0,774	0,914	0,115	2,12
	SNV	2,5	0,970	0,988	0,036	0,842	0,941	0,096	2,55
	Detrend	0,098	0,945	0,978	0,049	0,828	0,942	0,101	2,44
	SG	0,287	0,960	0,983	0,042	0,817	0,935	0,104	2,37
	1-Der	11,5	1,000	1,000	0,000	0,890	0,956	0,081	3,05
	2-Der	3902,13	1,000	1,000	0,000	0,871	0,946	0,087	2,82
	MSC	3,5	0,993	0,997	0,017	0,837	0,933	0,098	2,51
	DOSC	$1,39 \times 10^{10}$	1,000	1,000	0,000	0,786	0,920	0,112	2,19
	SNV + 2-Der	0,010	0,996	0,998	0,014	0,888	0,948	0,081	3,02
	MSC+2- Der	199,6	1,000	1,000	0,000	0,882	0,944	0,083	2,95

Table 3. Prediction results of tannins content using the average of the six measuring points with all wavelengths by PLS, SVM and LS-SVM models for the half cut fruit set

MODEL	PRE-PROCESSING	*LV, γ*	Calibration			Prediction			
			R^2	ρ	RMSE	R^2	ρ	RMSE	RPD
PLSR	**RAW**	17	0,991	0,996	0,020	0,812	0,921	0,105	2,33
	SNV	17	0,991	0,996	0,020	0,802	0,912	0,108	2,28
	Detrend	16	0,991	0,995	0,020	0,795	0,917	0,120	2,24
	SG	19	0,991	0,995	0,020	0,783	0,907	0,113	2,17
	1-Der	12	0,991	0,996	0,020	0,860	0,951	0,091	2,70
	2-Der	13	0,993	0,996	0,018	0,829	0,922	0,100	2,45
	MSC	16	0,990	0,995	0,021	0,794	0,908	0,110	2,23
	DOSC	1	1,000	1,000	0,000	0,804	0,917	0,107	2,28
	SNV + 2-Der	13	0,991	0,996	0,020	0,809	0,909	0,106	2,32
	MSC+2- Der	13	0,991	0,996	0,020	0,809	0,909	0,106	2,32
SVM	**RAW**		0,824	0,927	0,088	0,704	0,916	0,132	1,86
	SNV		0,972	0,987	0,035	0,862	0,939	0,090	2,72
	Detrend		0,939	0,970	0,052	0,864	0,942	0,090	2,74
	SG		0,761	0,888	0,102	0,625	0,876	0,148	1,65
	1-Der		0,966	0,985	0,038	0,881	0,957	0,084	2,93
	2-Der		0,979	0,991	0,030	0,891	0,956	0,080	3,07
	MSC		0,964	0,983	0,040	0,868	0,938	0,088	2,78
	DOSC		0,992	1,000	0,019	0,817	0,914	0,104	2,37
	SNV + 2-Der		0,975	0,990	0,033	0,888	0,952	0,081	3,03
	MSC+2- Der		0,979	0,991	0,031	0,888	0,953	0,081	3,03
LS-SVM	**RAW**	7,8	0,995	0,997	0,016	0,809	0,921	0,106	2,31
	SNV	0,039	0,993	0,997	0,017	0,844	0,930	0,096	2,56
	Detrend	0,057	0,987	0,995	0,024	0,835	0,924	0,098	2,49
	SG	1,8	0,984	0,993	0,026	0,805	0,916	0,107	2,29
	1-Der	0,035	0,999	0,999	0,003	0,883	0,951	0,083	2,95
	2-Der	36,6	1,000	1,000	0,000	0,893	0,953	0,079	3,09
	MSC	0,019	0,983	0,993	0,027	0,849	0,936	0,094	2,60
	DOSC	$8,83 \times 10^8$	1,000	1,000	0,000	0,804	0,917	0,107	2,28
	SNV + 2-Der	2,5	1,000	1,000	0,000	0,888	0,949	0,081	3,02
	MSC+2- Der	13,9	1,000	1,000	0,000	0,887	0,948	0,081	3,01

Table 4. Prediction results of tannins content using the average of the six measuring points with EWs (by SPA) by PLS, SVM and LS-SVM models for the intact fruit set

MODEL	PRE-PROCESSING	*EW/LV, EW, EW/γ*	Calibration			Prediction			
			R^2	ρ	RMSE	R^2	ρ	RMSE	RPD
PLSR	2-Der	16/16	0,861	0,928	0,078	0,853	0,924	0,093	2,64
	SNV + 2-Der	9/9	0,742	0,861	0,106	0,852	0,931	0,093	2,63
	MSC+2- Der	8/8	0,718	0,847	0,111	0,829	0,919	0,100	2,45
SVM	2-Der	16	0,841	0,917	0,084	0,893	0,946	0,079	3,10
	SNV + 2-Der	9	0,697	0,841	0,115	0,867	0,936	0,088	2,78
	MSC+2- Der	8	0,705	0,840	0,114	0,857	0,931	0,092	2,68
LS-SVM	2-Der	16/0,007	0,996	0,998	0,014	0,874	0,941	0,086	2,85
	SNV + 2-Der	9/0,005	0,992	0,997	0,018	0,882	0,947	0,083	2,95
	MSC+2- Der	8/0,003	0,985	0,994	0,025	0,881	0,948	0,084	2,94

Table 5. Prediction results of tannins content using the average of the six measuring points with EWs (by SPA) by PLS, SVM and LS-SVM models for the half cut fruit set

MODEL	PRE-PROCESSING	*EW/LV, EW, EW/γ*	Calibration			Prediction			
			R^2	ρ	RMSE	R^2	ρ	RMSE	RPD
PLSR	2-Der	18/12	0,786	0,887	0,097	0,771	0,878	0,116	2,12
	SNV + 2-Der	18/16	0,857	0,926	0,079	0,789	0,889	0,111	2,20
	MSC+2- Der	18/16	0,856	0,925	0,080	0,790	0,890	0,111	2,21
SVM	2-Der	18	0,739	0,860	0,107	0,760	0,877	0,119	2,07
	SNV + 2-Der	18	0,834	0,914	0,085	0,804	0,897	0,107	2,29
	MSC+2- Der	18	0,836	0,915	0,085	0,810	0,901	0,106	2,32
LS-SVM	2-Der	18/0,210	0,819	0,907	0,089	0,872	0,937	0,087	2,83
	SNV + 2-Der	18/0,628	0,725	0,852	0,120	0,871	0,942	0,087	2,82
	MSC+2- Der	18/8,2	0,717	0,847	0,111	0,836	0,923	0,098	2,50

Table 6. Selected effective wavelengths (EWs) from the average of the six measuring points by SPA for the intact fruit set and the half cut fruit set

DATA SET	PRE-PROCESSING	Nº	SELECTED EWs (nm)
Intact	2-Der	16	892.46, 896.38, 888.53, 691.26, 1781.94, 1791.80, 1079.55, 1051.26, 1073.71, 1068.98, 904.22, 1788.51, 1064.69, 1099.98, 1057.65, 1062.42
	SNV + 2-Der	9	892.46, 888.53, 691.26, 896.38, 1791.80, 1781.94, 1068.98, 1099.98, 1064.69
	MSC+2- Der	8	892.46, 888.53, 691.26, 896.38, 1791.80, 1781.94, 1068.98, 1064.69
Half-cut	2-Der	18	1791.80, 896.38, 1099.54, 1031.00, 1063.33, 1099.76, 1071.24, 892.46, 1099.32, 888.53, 1067.85, 1060.83, 594.85, 1079.10, 1068.98, 1099.98, 1066.95, 1054.91
	SNV + 2-Der & MSC+2- Der	18	888.53, 892.46, 1791.8, 1100.20, 896.38, 1065.14, 904.22, 846.10, 1070.78, 1072.14, 1067.40, 1785.23, 595.69, 1075.51, 1068.98, 1054.91, 1099.32, 1071.24

Table 7. Prediction results of tannins content using different combinations of measuring points and pre-processing with all wavelengths by PLS, SVM and LS-SVM models*

POINTS	MODEL	Entire					Half cut				
		BEST PRE-TREATMENT	R^2	ρ	RMSE	RPD	BEST PRE-TREATMENT	R^2	ρ	RMSE	RPD
1-6-2-5	PLSR	MSC+ 2-Der	0,857	0,927	0,092	2,68	1-Der	0,859	0,952	0,091	2,69
	SVM	MSC+2-Der	0,890	0,952	0,080	3,05	MSC+2-Der	0,875	0,947	0,086	2,86
	LS-SVM	SNV+ 2-Der	0,882	0,941	0,083	2,95	2-Der	0,865	0,937	0,089	2,76
1-6-3-4	PLSR	SNV+2-Der	0,844	0,929	0,096	2,56	MSC+2-Der	0,865	0,934	0,089	2,76
	SVM	MSC+2-Der	0,887	0,957	0,081	3,01	2-Der	0,888	0,960	0,081	3,02
	LS-SVM	SNV+ 2-Der	0,876	0,950	0,085	2,87	SNV+2-Der	0,888	0,950	0,081	3,03
1-6	PLSR	1-Der	0,790	0,916	0,111	2,21	2-Der	0,827	0,927	0,101	2,43
	SVM	MSC+2-Der	0,850	0,941	0,094	2,61	MSC+2-Der	0,865	0,945	0,089	2,75
	LS-SVM	1-Der	0,836	0,937	0,098	2,50	MSC+2-Der	0,859	0,935	0,091	2,69
2-5	PLSR	1-Der	0,773	0,886	0,116	2,12	Detrend	0,811	0,917	0,105	2,33
	SVM	2-Der	0,871	0,944	0,087	2,82	MSC	0,839	0,937	0,097	2,52
	LS-SVM	2-Der	0,860	0,931	0,091	2,70	1-Der	0,845	0,943	0,096	2,57
3-4	PLSR	1-Der	0,798	0,912	0,109	2,25	2-Der	0,845	0,921	0,095	2,57
	SVM	MSC+2-Der	0,875	0,955	0,086	2,86	SNV+2-Der	0,842	0,949	0,096	2,55
	LS-SVM	SNV+2-Der	0,862	0,949	0,090	2,73	SNV+2-Der	0,855	0,942	0,092	2,66

* Only the best prediction results for each model are shown, indicating the associated pre-processing

Table 8. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with all wavelengths by PLS, SVM and LS-SVM models for the intact fruit set

MODEL	PRE-PROCESSING	*LV*, γ	Calibration			Prediction			
			R^2	ρ	RMSE	R^2	ρ	RMSE	RPD
PLSR	**RAW**	19	0,992	0,996	0,019	0,768	0,890	0,117	2,10
	SNV	18	0,991	0,995	0,020	0,799	0,904	0,109	2,26
	Detrend	18	0,992	0,996	0,019	0,771	0,890	0,116	2,11
	SG	21	0,991	0,996	0,0199	0,750	0,875	0,121	2,02
	1-Der	10	0,990	0,995	0,021	0,797	0,907	0,109	2,24
	2-Der	12	0,990	0,995	0,021	0,810	0,905	0,106	2,32
	MSC	18	0,992	0,996	0,018	0,806	0,907	0,107	2,30
	DOSC	1	1,000	1,000	0,000	0,742	0,876	0,123	1,99
	SNV + 2-Der	12	0,991	0,995	0,020	0,825	0,911	0,1014	2,42
	MSC+2- Der	12	0,990	0,995	0,021	0,828	0,913	0,101	2,44
SVM	**RAW**		0,796	0,912	0,095	0,767	0,903	0,117	2,10
	SNV		0,942	0,971	0,050	0,851	0,931	0,094	2,62
	Detrend		0,948	0,974	0,048	0,837	0,922	0,098	2,51
	SG		0,767	0,886	0,101	0,729	0,888	0,126	1,94
	1-Der		0,978	0,991	0,031	0,875	0,944	0,086	2,86
	2-Der		0,981	0,992	0,029	0,885	0,944	0,082	2,98
	MSC		0,962	0,981	0,041	0,864	0,935	0,089	2,74
	DOSC		0,994	1,000	0,017	0,756	0,877	0,120	2,05
	SNV + 2-Der		0,966	0,984	0,039	0,896	0,953	0,078	3,13
	MSC+2- Der		0,980	0,993	0,029	0,892	0,950	0,080	3,07
LS-	**RAW**	0,207	0,944	0,975	0,050	0,810	0,913	0,106	2,32
SVM	**SNV**	0,056	0,896	0,951	0,067	0,814	0,932	0,105	2,35
	Detrend	0,016	0,949	0,978	0,047	0,827	0,922	0,101	2,43
	SG	0,181	0,916	0,962	0,061	0,804	0,912	0,107	2,29
	1-Der	2,2	1,000	1,000	0,000	0,854	0,937	0,093	2,65
	2-Der	0,007	0,996	0,998	0,014	0,874	0,941	0,086	2,85
	MSC	511,1	0,910	0,957	0,063	0,826	0,933	0,101	2,43
	DOSC	$9.45x10^9$	1,000	1,000	0,000	0,742	0,876	0,123	1,99
	SNV + 2-Der	0,005	0,992	0,997	0,018	0,882	0,947	0,083	2,95
	MSC+2- Der	0,003	0,985	0,994	0,025	0,881	0,948	0,084	2,94

Table 9. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with all wavelengths by PLS, SVM and LS-SVM models for the half cut fruit set

MODEL	PRE-PROCESSING	*LV*, γ	Calibration			Prediction			
			R^2	ρ	RMSE	R^2	ρ	RMSE	RPD
PLSR	**RAW**	17	0,991	0,996	0,020	0,803	0,924	0,108	2,28
	SNV	17	0,992	0,996	0,019	0,785	0,914	0,112	2,18
	Detrend	16	0,990	0,995	0,021	0,783	0,917	0,113	2,17
	SG	19	0,991	0,996	0,020	0,790	0,918	0,111	2,21
	1-Der	11	0,990	0,995	0,021	0,842	0,942	0,096	2,55
	2-Der	12	0,991	0,996	0,020	0,838	0,929	0,098	2,52
	MSC	16	0,992	0,996	0,019	0,793	0,916	0,110	2,22
	DOSC	1	1,000	1,000	0,000	0,786	0,920	0,112	2,19
	SNV + 2-Der	12	0,991	0,996	0,020	0,819	0,915	0,103	2,38
	MSC+2- Der	12	0,991	0,996	0,020	0,818	0,915	0,103	2,38
SVM	**RAW**		0,757	0,878	0,103	0,699	0,889	0,133	1,84
	SNV		0,964	0,983	0,040	0,866	0,943	0,089	2,77
	Detrend		0,962	0,982	0,041	0,849	0,934	0,094	2,60
	SG		0,736	0,867	0,108	0,600	0,845	0,153	1,60
	1-Der		0,976	0,990	0,033	0,877	0,960	0,085	2,89
	2-Der		0,979	0,991	0,031	0,859	0,947	0,091	2,70
	MSC		0,947	0,974	0,048	0,864	0,946	0,090	2,74
	DOSC		0,992	1,000	0,019	0,797	0,911	0,109	2,25
	SNV + 2-Der		0,979	0,992	0,031	0,859	0,945	0,091	2,70
	MSC+2- Der		0,979	0,992	0,030	0,858	0,946	0,092	2,68
LS-SVM	**RAW**	5651,36	0,944	0,974	0,050	0,774	0,914	0,115	2,13
	SNV	0,011	0,970	0,988	0,036	0,842	0,941	0,096	2,55
	Detrend	0,014	0,945	0,978	0,049	0,828	0,942	0,101	2,44
	SG	0,704	0,960	0,983	0,042	0,817	0,935	0,104	2,37
	1-Der	1722,6	1,000	1,000	0,000	0,890	0,956	0,081	3,05
	2-Der	651,6	1,000	1,000	0,000	0,871	0,946	0,087	2,82
	MSC	0,036	0,993	0,997	0,017	0,837	0,933	0,098	2,51
	DOSC	4.57x107	1,000	1,000	0,000	0,786	0,920	0,112	2,19
	SNV + 2-Der	8,4	1,000	1,000	0,000	0,867	0,943	0,089	2,77
	MSC+2- Der	37,9	1,000	1,000	0,000	0,866	0,943	0,089	2,77

Table 10. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with EWs (selected by SPA) by PLS, SVM and LS-SVM models for the intact fruit set

MODEL	PRE-PROCESSING	*EW/LV, EW, EW/γ*	Calibration			Prediction			
			R^2	ρ	RMSE	R^2	ρ	RMSE	RPD
PLSR	**2-Der**	22/15	0,849	0,921	0,081	0,828	0,912	0,101	2,44
	SNV +2-Der	20/11	0,809	0,899	0,092	0,787	0,890	0,112	2,19
	MSC+2-Der	13/11	0,780	0,883	0,098	0,815	0,910	0,104	2,36
SVM	**2-Der**	22	0,830	0,912	0,086	0,819	0,905	0,103	2,38
	SNV +2-Der	20	0,746	0,867	0,106	0,845	0,921	0,096	2,57
	MSC+2-Der	13	0,744	0,865	0,106	0,776	0,890	0,115	2,13
LS-SVM	**2-Der**	22/1,46	0,844	0,919	0,083	0,819	0,906	0,103	2,38
	SNV +2-Der	20/4,7	0,807	0,899	0,092	0,806	0,902	0,107	2,30
	MSC+2-Der	13/0,201	0,734	0,868	0,108	0,681	0,879	0,137	1,79

Table 11. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with EWs (selected by SPA) by PLS, SVM and LS-SVM models for the half cut fruit set

MODEL	PRE-PROCESSING	*EW/LV, EW, EW/γ*	Calibration			Prediction			
			R^2	ρ	RMSE	R^2	ρ	RMSE	RPD
PLSR	**1-Der**	14	0,951	0,975	0,047	0,771	0,893	0,116	2,11
	2-Der	18	0,919	0,959	0,060	0,738	0,861	0,124	1,98
	SNV+2-Der	26	0,889	0,943	0,070	0,779	0,895	0,114	2,15
SVM	**1-Der**	37	0,912	0,955	0,062	0,763	0,879	0,118	2,08
	2-Der	39	0,846	0,920	0,082	0,742	0,864	0,123	1,99
	SNV+2-Der	37	0,839	0,918	0,084	0,807	0,923	0,106	2,31
LS-SVM	**1-Der**	37/ $9.89x10^7$	0,952	0,976	0,046	0,774	0,894	0,115	2,13
	2-Der	39 / 3,7	0,920	0,959	0,059	0,757	0,872	0,120	2,05
	SNV+2-Der	37 / 0,933	0,882	0,939	0,072	0,802	0,906	0,108	2,27

Table 12. Selected effective wavelengths (EWs) from the average of the four measuring points (2-5-3-4) by SPA for the intact fruit set and the half cut fruit set

DATA SET	PRE-PROCESSING	Nº	SELECTED EWs (nm)
Intact	**2-Der**	22	892.46, 1781.94, 1791.80, 1099.54, 888.53, 1099.76, 1050.35, 929.01, 1058.33, 1064.69, 1785.23, 1051.26, 1068.08, 1073.71, 1099.32, 1069.43, 1079.55, 1065.14, 594.85, 1072.81, 1059.47, 1055.37
	SNV + 2-Der	20	892.46, 896.38, 1791.80, 1099.54, 1099.76, 904.22, 888.53, 1099.32, 1058.33, 1059.24, 1065.14, 1781.94, 1072.36, 1064.69, 1062.42, 1069.43, 1099.98, 1054.00, 1072.81, 1077.53
	MSC+2- Der	13	1058.33, 1064.69, 888.53, 1073.71, 1791.80, 1053.32, 1059.24, 1781.94, 689.62, 1099.54, 1054.00, 1079.55, 1066.50
Half-cut	**1-Der**	37	1100.20, 888.53, 892.46, 1788.51, 1099.76, 1099.32, 900.30, 1795.09, 1087.36, 1071.01, 1099.98, 1079.32, 1043.93, 1084.24, 1067.17, 1099.54, 1077.08, 1064.91, 1081.56, 1068.98, 1079.55, 1070.33, 1081.11, 1049.89, 896.38, 1077.75, 1085.58, 1082.23, 1053.09, 1082.90, 1071.24, 1054.23, 1070.78, 1057.42, 1075.51, 1076.63, 1065.36
	2-Der	39	1791.80, 896.38, 888.53, 1072.14, 1070.11, 1070.78, 892.46, 1058.10, 1068.30, 1068.75, 1077.98, 1074.16, 1084.02, 1077.53, 1065.14, 1057.19, 1073.71, 1082.68, 1099.32, 1785.23, 1071.46, 1074.84, 1099.54, 594.85, 1056.51, 1781.94, 1061.06, 1065.59, 1055.83, 1081.78, 1067.85, 1072.81, 1061.51, 1060.60, 1063.78, 660.46, 1079.10, 1076.86, 1074.39
	SNV+ 2-Der	37	888.53, 1100.20, 1791.80, 896.38, 1486.16, 1068.30, 1099.98, 1070.11, 892.46, 1065.14, 1082.68, 1068.98, 1788.51, 1077.30, 1061.06, 1072.81, 1072.14, 1061.51, 1060.60, 1058.10, 1078.20, 1074.39, 1064.23, 1078.65, 1067.40, 1054.91, 1080.44, 1059.92, 904.22, 1075.06, 1066.95, 1072.36, 1063.78, 1075.51, 1069.66, 1781.94, 1081.78