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

Victoria Cortés[1,3], Alejandro Rodríguez[2], José Blasco[3], Beatriz Rey[2], Cristina Besada[4], Sergio Cubero[3], Alejandra Salvador[4], Pau Talens[1], 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 control of fruit quality requires reliable and rapid determination techniques. Therefore, the food industry has a growing interest in non-destructive methods such as spectroscopy. The aim of this study was to evaluate the feasibility of visible and near-infrared (NIR) spectroscopy, in combination with multivariate analysis techniques, to predict the level and changes of astringency in intact and in the flesh of half cut persimmon fruits. The fruits were harvested and exposed to different treatments with 95 % CO_2 at 20 ºC for 0, 6, 12, 18 and 24 h to obtain samples with different levels of astringency. A set of 98 fruits was used to develop the predictive models based on their spectral data and another external set of 42 fruit samples was used to validate the models. The models were created using the partial least squares regression (PLSR), support vector machine (SVM) and least squares support vector machine (LS-SVM). In general, the models with the best performance were those which included standard normal variate (SNV) in the pre-processing. The best model was the PLSR developed with SNV along with the first derivative (1-Der) pre-processing, created using the data obtained at six measurement points of the intact fruits and all wavelengths (R^2=0.904 and RPD=3.26). Later, a successive projection algorithm (SPA) was applied to select the most effective wavelengths (EWs). Using the six points of measurement of the

intact fruit and SNV together with the direct orthogonal signal correction (DOSC) pre-processing in the NIR spectra, 41 EWs were selected, achieving an R^2 of 0.915 and an RPD of 3.46 for the PLSR model. These results suggest that this technology has potential for use as a feasible and cost-effective method for the non-destructive determination of astringency in persimmon fruits.

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

1. INTRODUCTION

Persimmon (*Diospyros kaki* L.) is a fruit originally from China, but is now cultivated in warm regions around the world (Ashtiani et al., 2016). The climatic characteristics of the production are important factors that influence the quality and properties of the fruits. The main areas where this fruit is cultivated in Spain are Alicante, Andalucía, Castellón, Extremadura and Valencia, especially in Ribera del Xúquer, which was granted Protected Designation of Origin (PDO) status by the Spanish government in 1998 (Khanmohammadi et al., 2014). Several cultivars of persimmon are grown in Spain, such as the astringent type 'Rojo Brillante'. Persimmon develops an astringent taste due to the presence of soluble tannins. Tannins are polyphenol compounds with a high molecular weight and their large hydroxyl phenolic groups cause astringency. As the fruit ripens, the soluble tannins gradually turn into insoluble tannins, making the fruit less astringent (Noypitak et al., 2015). However, several postharvest treatments can be applied to achieve the fast removal of the astringency of the fruits without affecting the firmness of the pulp (Khademi et al., 2010). Among them, the most widely used commercial technique is based on exposing the fruits to a high concentration of CO_2 (95%–98%). This method promotes anaerobic respiration in the fruit, resulting in an accumulation of acetaldehyde, which reacts with the soluble tannins. The tannins become insoluble with the treatment and the astringency is thus eliminated (Matsuo et al., 1991). If the treatment is too short, it can result in fruits with residual astringency (Besada et al., 2010), whereas if it is too long, it may lead to loss of fruit quality (Novillo et al., 2014). Therefore, it is important to investigate non-destructive techniques to ensure the success of the treatments.

Techniques based on the spectrum analysis, like hyperspectral imaging, have been widely used for the qualitative and quantitative determination of different properties in

fruit (Lorente et al., 2012). Munera et al. (2017b & 2017a) analysed the astringency and the internal quality of persimmon using hyperspectral imaging, which has the advantage of obtaining both spectral and spatial information. However, one of the most common techniques currently used in food chemistry is near-infrared (NIR) spectroscopy, as it is non-destructive, inexpensive, rapid and reliable (Nicolaï et al., 2007; Vitale et al., 2013; López et al., 2013). This technique has been used for the quantitative determination of several internal properties or compounds (Schmilovitch et al., 2000; Nagle et al., 2010; Theanjumpol et al., 2013), to determine maturity (Jha et al., 2012) and also to measure quality indices (Attila & János, 2011; Cortés et al., 2016).

The combination of chemometrics and spectroscopy has been applied in the food industry, agriculture and horticulture to obtain information from spectra. Support vector machine (SVM) are learning algorithms used for classification and regression tasks widely used in the analysis of spectroscopic data (Devos et al., 2009; Fernadez-Pierna et al., 2012). Chauchard et al. (2004) compared classical linear regression techniques with least square-support vector machine (LS-SVM) regression to predict the total acidity in fresh grapes using NIR spectroscopy. LS-SVM in combination with Standard normal variate (SNV) pre-processing and partial least square regression (PLSR) latent variables increased the rate of prediction. Nicolaï et al. (2007) predicted sugar content using PLSR. The covariance, Gaussian and cubic polynomial kernel functions obtained similar results of about R^2=0.87 and Q^2=0.84 for all methods, concluding that kernel PLSR offered no advantages compared to ordinary PLSR. The identification of the spectral variables (wavelengths) can lead to better classification results and simplify the chemical interpretation of the results. Calvini et al. (2015) tested sparse principal component analysis (PCA) together with k-Nearest-Neighbours (k-NN) and sparse PLS discriminant analysis (PLS-DA) to discriminate between Arabica and Robusta coffee, and compared the results with the classical approaches based on PCA+kNN and PLS-DA.

Lorente et al. (2015) used NIR spectroscopy (650 to 1700 nm) to detect early invisible decay lesions in citrus fruit using MSC and SNV pre-processing, different methods to select the important bands, and linear discriminant analysis (LDA) to classify the fruit as being either sound or rotten with a rate of correct classification above 90 %. Folch-Fortuny et al. (2016) used N-way-PLS-DA to detect early invisible decay lesions in citrus fruit, achieving a prediction rate higher than 90 %. Mowat and Poole (1997) found this technology useful in determining persimmon quality. Ito et al. (1997) and

Noypitak et al. (2015) investigated astringency in the persimmons 'Nisimura-wase' and in 'Xichu', respectively. The most common mode used in NIRS is diffuse reflectance, which acquires the reflected light in the vicinity of the illuminating point and is preferable for the measurement of intact fruit (Shao et al., 2009; He et al., 2007).

The aim of this study was to evaluate the feasibility of visible and NIR spectroscopy combined with chemometrics as a non-destructive tool to determine 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 stages of commercial maturity (M1 and M2) corresponding to late November and mid-December. The maturity index used to select the fruits was a visual observation of the external colour of the fruit (Salvador et al., 2007). After each harvest, 70 fruits without external damage and of homogenous colour were selected (a total of 140 fruits). In order to characterise the fruit at harvest, the average colour index (CI=100*a*/*Lb*, Hunter parameters) was measured using a colorimeter (CR-300, Konica Minolta Inc, Tokyo, Japan) and the firmness of the flesh was measured by a universal testing machine (4301, Instron Engineering Corp., MA, USA) equipped with an 8 mm puncture probe. The crosshead speed during the firmness test was 10 mm/min. During the test, the force increased slowly until it decreased abruptly when the flesh broke, and then the maximum required force (in Newton) was recorded.

The average CI resulted in 18.20 ± 3.32 for M1 and 21.6 ± 4.05 for M2, while firmness decreased along with maturity at harvest, with mean values being 30.8 N ± 3.5 and 24.4 N ± 4.9 for M1 and M2, respectively.

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

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

The degree of astringency of each fruit was determined as follows. A flesh sample of each fruit was frozen at −20 ºC and the soluble tannin content was analysed using the Folin-Denis method (Taira, 1995). The results were expressed as relative soluble tannins by fresh weight. Prior to this process, each fruit was cut in half and pressed onto 10x10 cm filter paper soaked in a solution of 5 % $FeCl_3$, which resulted in an impression whose quantity and intensity gave information about the content of soluble tannins and their distribution (Matsuo & Ito, 1982). This method of tannin printing is an alternative technique to the Folin-Denis method used in industry in random fruits to determine the level of astringency in fruit lots.

3. Visible and near-infrared spectra collection

The spectra were alternately collected in reflectance mode using a multi-channel 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 working visible and near-infrared (VNIR) range from 650 nm to 1050 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 working NIR range from 1000 nm to 1700 nm with 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. The probe tip was designed to provide reflectance measurements at a 45º angle so as to minimise the specular reflectance of the fruit surface.

Calibration was performed using a 99 % white reflective reference tile (WS-2, Avantes BV, The Netherlands) so that the maximum reflectance of the reference measured over the entire spectral range was 90 % of the value of saturation. Before taking the spectral measurements, the temperature of the persimmons was stabilised at 24 ºC. Measurements were performed at the six different points on the surface of the intact persimmon and the flesh of half cut fruit (Fig. 1), and mean 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 and pre-process the spectra. The integration time was set at 90 ms for the detector sensitive in the VNIR and 700 ms for the detector sensitive in the NIR region. For both detectors, each spectrum was obtained as the average of five scans in order to reduce the detector's thermal noise (Nicolaï et al., 2007). The mean reflectance measurements of each sample (S) were then converted to relative reflectance (R) values with respect to the white reference using dark reflectance (D) values 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 switching off the light source and covering the whole tip of the reflectance probe.

Figure 2. A labelled picture of the spectrometer

4. Statistical analysis

Spectral data and the tannin reference values were organised into matrices, where the rows represented the samples (the total of 140 persimmons) and the columns represented the variables. The X-variables, or predictors, were the wavelengths of the VNIR and NIR spectra for each persimmon. The Y-variable, or response, in the last column, represented the measured tannin value associated with each sample.

A total of 28 matrices were generated corresponding to different combinations of the measurement points of the intact fruit and the flesh of the half cut fruit. The first two matrices corresponded to the mean values of reflectance of the measurements at the six points of the intact fruit shown in Figure 1. The third and fourth matrices contained mean values of the measurements at four points (2-5-3-4), which corresponded to the lowest part of the intact persimmon in the VNIR and NIR detectors, respectively. The fifth to fourteenth matrices contained mean values for measurements of other combinations of points (1-6-2-5, 1-6-3-4, 1-6, 2-5 and 3-4) in both VNIR and NIR. Other combinations of measured points have not been taken into account since the deastringency process normally progresses from the top to the bottom of the fruit (Fig. 5) and would not make sense. The remaining 14 matrices corresponded to the mean values of the measurements of the same combinations of points, but from the flesh of the half cut fruit.

4.1. Spectral Pre-processing

To remove the influence of unwanted effects such as high-frequency noise, baseline shifts, light scattering, random noise and any other external effects due to instrumental or environmental factors, six methods of spectral pre-processing and their combinations were applied before the development of the prediction models. These methods included standard normal variate (SNV), multiplicative scatter correction (MSC), 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 prediction models were carried out using MATLAB R2015b (The Mathworks Inc., Natick, MA, USA).

SNV is commonly used to eliminate the multiplicative noise due to the influence of particle size or scatter interference (Rinnan et al., 2009). SNV subtracts the mean from an individual spectrum and divides it by its standard deviation (Feng & Sun, 2013). Similarly, MSC is used to compensate for the non-uniform scattering effect induced by diverse particle sizes and other physical effects in the spectrum (Fearn et al., 2009; Vidal & Amigo, 2012). It linearises each spectrum to an average spectrum (derived from the calibration set) and adjusts it using the least squares method.

Moreover, smoothing is an effective way to reduce high-frequency noise. There are several smoothing methods in the literature, but one of the most commonly applied is SG smoothing (Savitzky & Golay, 1964). This method has the advantage of preserving signal characteristics such as the maximum and minimum relative values or the width of the peaks, which usually disappear with other smoothing methods. In the present work, SG smoothing was calculated with two-degree polynomials and a window size of seven points.

1-Der and 2-Der are well-accepted pre-processing methods to eliminate the shifting, the scattering and the background noise, as well as to distinguish overlapping peaks and to improve the spectral resolution (Sinija & Mishra, 2011). They were calculated using the SG algorithm with three-point smoothing filters and a two-degree polynomial (Liu et al., 2010).

Finally, DOSC are novel methods used to remove information that has a poor correlation (orthogonal) with the response matrix (Zhu et al., 2008). DOSC obtains components that are orthogonal to the response matrix and eliminates those that are considered irrelevant, thus improving the predictability.

4.2. Modelling by different calibration methods

Estimation of prediction error is required to evaluate the performance of fitted models. Cross-validation is widely used to estimate the prediction error (Fusiki, 2011). In this work, 70 % of the fruits in each maturity stage were randomly selected to build the models that were internally validated using a 10-fold cross-validation. The remaining 30 % of the samples were never used to build or train the model with the purpose of externally evaluating the performance of the regression techniques used to predict the tannin content. The regression techniques used in this work were PLSR, SVM and the LS-SVM regression.

The PLSR multivariate method is widely used to evaluate the linear relationship between inputs (spectral data or X-variables) and the response variable (tannin content in this case or Y-variable) in spectroscopic analysis (Geladi & Kowalski, 1986). 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 (Lorber et al., 1987). SVM is a popular machine learning tool for regression (Vapnik, 2013) based on the Vapnik-Chervonenkis (VC) dimension and on the principle of structural risk minimisation (Gunn, 1998). It is considered a non-parametric technique because the SVM models are based on a non-linear kernel function. In short, SVM assigns the calibration dataset to a high-dimensional feature space by means of a non-linear mapping, and then performs a linear regression. This technique has the advantage of being very efficient and robust during the training of the model. In this study, the Matlab statistical and machine learning toolbox was used to train the model with the spectral and tannin information, using a linear kernel and a 10-fold cross-validation.

Finally, LS-SVM is a learning algorithm which improves the generalisation ability of the machine learning procedure based on the principle of structural risk minimisation (Liu et al., 2008; Suykens & Vandewalle, 1999). It handles both linear and non-linear 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 (Liu & Sun, 2009). The LS-SVM has the advantage of limited over-fitting, high predictive reliability and a strong generalisation capability. The LS-SVMlab v1.8 toolbox (Suykens, Leuven, Belgium) was used to develop the calibration models. During the development of the model, the linear kernel

and a 10-fold cross-validation were used to avoid problems of over-fitting. The linear kernel included a regularisation parameter that determined the trade-off between minimising the training error and minimising the model complexity. A large γ implies little regularisation, and therefore a more non-linear model (Sun et al., 2009).

4.3. Variable selection

Since the number of variables used as inputs (wavelengths) in the models is high (1570 variables for the VNIR and 198 for the NIR spectra), they may contain excessive collinearity and redundancy. Therefore, it was considered appropriate to find the most important wavelengths as effective wavelengths (EWs) for each model. This was performed 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). SPA is a variable selection algorithm applied to solve collinearity problems and to select the wavelengths with fewer redundancies by means of a simple procedure of projection in a vectorial space, thereby allowing for the selection of the best subsets of wavelengths that conform to the minimum collinearity (Araújo et al., 2001; Galvao et al., 2008; Zhang et al., 2013). SPA was applied for each calibration set and the EWs obtained were used again as inputs of the PLSR, SVM and LS-SVM models.

4.4. Model evaluation

The accuracy and the predictive capability of the three different models were evaluated by means of the coefficient of determination (R^2), the root mean square error (RMSE) and the ratio of performance to deviation (RPD) obtained on the external validation set. Generally, a good model must have high R^2 with low RMSE. In addition, an acceptable model should have an RPD value of more than 2.5, a value above 3.0 being very good (Williams & Sobering, 1993; Viscarra Rossel et al., 2007; Kamruzzaman et al., 2016; Cortés et al., 2016). These parameters can be defined by equations 2 to 4.

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

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

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

where:

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

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

N: is the number of observations.

SD: is the standard deviation of the measured values.

5. RESULTS AND DISCUSSION

The total number of persimmon samples was 140, with a mean tannin content of 0.250 % (STD=0.221). The statistical values of the persimmon tannin content in the calibration and external validation sets are shown in Table 1.

Table 1. Statistical values of tannin content (%) of the studied persimmons

Before applying the models, the raw reflectance spectra (Fig. 3) of the samples were pre-processed using the described methods.

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

Thus, the PLSR, SVM and LS-SVM models were developed using both raw and pre-processed spectra. Samples in the external validation set were later used to evaluate the performance of the models. The results (R^2, RMSE and RPD) of the models for the external validation set are shown in Table 2 and Table 3. Table 2 shows the results using the average of the six measurement points for the intact fruit set, and Table 3 for the half cut fruit set.

Table 2. Results of tannin content using the average of the six measurement points with all wavelengths by PLSR, SVM and LS-SVM models for the intact fruit set

Table 3. Results of tannin content using the average of the six measurement points with all wavelengths by PLSR, SVM and LS-SVM models for the half cut fruit set

Tables 2 and 3 show that, on average, the models with the best performance are those which included SNV in the pre-processing that was applied (SNV+1-Der, SNV+2-Der,

SNV+DOSC). Figure 4 shows the results for the best PLSR model, which was obtained with the spectra measured at the six measurement points of the intact fruits and pre-processed using SNV+1-Der.

Figure 4. Normalised X-loading weights of the best PLSR model for the six measurement points (with SNV+1-Der pre-processing for the intact fruit set) for the (a) VNIR and (b) NIR detectors, respectively. Only the weights corresponding to the latent variables that explain 95 % of the Y-variable variance are shown (5 for VNIR and 16 for NIR detectors)

Tables 4 and 5 show the results for the three selected methods and the above mentioned pre-processing combinations after applying SPA for wavelength selection.

Table 4. Results of tannin content using the average of the six measurement points with EWs for the models created by PLSR, SVM and LS-SVM for the intact fruits set

Table 5. Results of tannin content using the average of the six measurement points with EWs for the models created by PLSR, SVM and LS-SVM for the half cut fruit set

This analysis was performed for the different combinations of the six measurement points, obtaining the results in Table 6, which shows the best results for each combination of points and each model. Tables 7 and 8 show the results for the combination of measurement points 2-5-3-4 (average of the equator and bottom of the fruit) for the intact and half cut fruit sets, respectively, for the three models (PLSR, SVM, LS-SVM), and the best pre-processing combinations for the six measurement points (SNV+1-Der, SNV+2-Der and SNV+DOSC). The highest RPD achieved was always equal to or better than the highest RPD obtained with any other combination of points. This is reasonable, since from the tannin prints observed in Figure 5, which were obtained using the technique based on $FeCl_3$, the highest differences are in the lower part of the fruit, the upper part being more similar in fruits with different CO_2 treatments (Fig. 5b-e).

Figure 5. Impressions of tannin content representing the evolution of the astringency distribution and intensity for persimmons after different CO_2 treatments: a) untreated; and b-e) treated with CO_2 for 6, 12, 18 and 24 h, respectively

371

372 **Table 6.** Results of tannin content achieved using different combinations of
373 measurement points and pre-processing methods with all wavelengths by PLSR, SVM
374 and LS-SVM models

375

376 **Table 7**. Results of tannin content achieved using the average of the four measurement
377 points (2-5-3-4) with all wavelengths by PLSR, SVM and LS-SVM models for the
378 intact fruit set

379

380 **Table 8**. Results of tannin content achieved using the average of the four measurement
381 points (2-5-3-4) with all wavelengths by PLSR, SVM and LS-SVM models for the half
382 cut fruit set

383 As in the previous case, SPA was applied for wavelength selection. Tables 9 and 10
384 show the results of these analyses for the three models and pre-processing
385 combinations.

386

387 **Table 9**. Results of tannin content achieved using the average of the four measurement
388 points (2-5-3-4) by PLSR, SVM and LS-SVM models with EWs selected by SPA for
389 the intact fruit set

390 **Table 10**. Results of tannin content achieved using the average of the four measurement
391 points (2-5-3-4) by PLSR, SVM and LS-SVM models with EWs selected by SPA for
392 the half cut fruit set

393 In this work, different models were obtained to estimate the content of tannins in
394 persimmon from their original and pre-processed reflectance spectra. The models were
395 created for measurements of the skin (intact fruit) and the flesh (half cut fruit). For the
396 intact fruit, good results were obtained for the three methods analysed (PLSR, SVM and
397 LS-SVM), achieving an RPD>3 in the best cases, using the average of the six
398 measurement points. The best results using the prediction set were obtained using PLSR
399 and SNV+1-Der pre-processing, in the VNIR region (RPD=3.26, R^2=0.904,
400 RMSE=0.075). Using SVM, the best results were for the NIR spectra and the
401 SNV+DOSC pre-processing. However, the analysis of the VNIR spectra using SVM
402 gave similar results with some pre-processing such as SNV+2-Der. Finally, the best

results with the LS-SVM method were obtained with the SNV+DOCS pre-processing in the NIR region. Regarding half cut fruit and the average of six measurement points, the results were poorer than in the case of intact fruit.

The selection of the most important wavelengths using SPA generally improves the results, especially in the case of half cut fruit. A model with an RPD greater than 3 was obtained for the VNIR spectra with the SNV+2-Der pre-processing and SVM method. In the case of the intact fruit, although the results did not always improve, the best result of the study was obtained using PLSR with SNV+DOSC in the NIR region, with an RPD of 3.46 (R^2=0.915, RMSE=0.071). As shown in Figure 4a, the values of the loading weights were higher around the 1000 nm band for the VNIR range, which corresponds to the information presented by Noypitak et al. (2015) in relation with the spectrum for the tannic acid powder. These loadings explained the better results obtained with the VNIR probe over those obtained in the NIR, and also the reduced number of EWs obtained in the VNIR range.

For both the intact and the half cut fruit cases, the three methods analysed achieved poorer predictions using the average of the four measurement points (combination 2-5-3-4) than those obtained with the six measurement points. Regarding the selection of EWs with SPA (with this combination of points), this method also improved the results obtained for the half cut fruit, similarly to the results obtained with six measurement points. However, the SPA analysis showed no significant improvement in intact fruit (RPD<3).

6. CONCLUSIONS

This study points to visible and NIR spectroscopy as a non-destructive method suitable for determining astringency in persimmon fruits in an easy and rapid way without expensive and tedious chemical analysis or the subjective evaluation of the tannin print method. Reflectance spectra at selected points in intact and half cut persimmons were acquired in the VIS and NIR regions. A total of seven signal pre-processing methods including SNV, SG, 1-Der, 2-Der, MSC, DOSC and combinations of them have been used in the measurements of the single point and the combination of selected points. The combinations considered were SNV+1-Der, SNV+2-Der and SNV+DOSC, since they showed the best performance from all the combinations evaluated. Astringency in persimmon fruits was predicted using three regression techniques, such as PLSR, SVM, and LS-SVM.

In addition, EWs were obtained using SPA. Depending on the method, the EWs varied from 1 to 30 when the VNIR spectra were used and from 17 to 57 when using the NIR spectra.

The best performance for intact fruits was obtained using PLSR on the full spectra of the six measurement points after pre-processing with SNV+1-Der, an R^2=0.904 and RPD=3.26 being achieved. Moreover, the best prediction results obtained with the EWs (41 bands) were obtained for the PLSR model using the six measurement points of the intact fruit in the NIR spectra and SNV+DOSC pre-processing (R^2=0.915; RPD=3.46). Hence, this technology has proved itself to be a feasible non-destructive method to determine the astringency in persimmon fruits, since the best results were achieved in intact fruits.

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-C04-01/03, RTA2013-00043-C02, and RTA2015-00078-00-00 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

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

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

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

Beghi, R., Giovanelli, G., Malegori, C., Giovenzana, V. & Guidetti, R. (2014). Testing of a VIS-NIR System for the Monitoring of Long-Term Apple Storage. Food and Bioprocess Technology, 7, 2134–2143.

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

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

Chauchard, F., Cogdill, R., Roussel, S., Roger, J.M., & Bellon-Maurel, V. (2004). 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, 71, 141-150.

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.

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

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

Feng, Y. Z., & Sun, D. W. (2013). 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, 109, 74-83.

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

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

Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, 21, 137–146

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

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

508 Gunn, S. R. (1998). Support vector machines for classification and regression. ISIS
509 technical report, 14.

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

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

515 Jha, S.N., Jaiswal P, Narsaiah K, Sharma R, Bhardwaj R, Gupta M, & Kumar R. (2013).
516 Authentication of mango varieties using near infrared spectroscopy. Agricultural
517 Research, 2(3), 229–235.

518 Kamruzzaman, M., Makino, Y., & Oshita, S. (2016). Rapid and non-destructive
519 detection of chicken adulteration in minced beef using visible near-infrared
520 hyperspectral imaging and machine learning. Journal of Food Engineering, 170, 8-15.

521 Khademi, O., Mostofi, Y., Zamani, Z. & Fatahi, R. (2010). The effect of deastringency
522 treatments on increasing the marketability of persimmon fruit. Acta Horticulturae, 877,
523 687–691.

524 Khanmohammadi, M., Karami, F., Mir-Marqués, A., Bagheri Garmarudi, A., Garrigues,
525 S. & de la Guardia, M. (2014). Classification of persimmon fruit origin by near infrared
526 spectrometry and least squares-support vector machines. Journal of Food Engineering,
527 142, 17-22.

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

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

534 Liu, Y., Sun, X., Zhou, J., Zhang, H., & Yang, C. (2010). Linear and nonlinear
535 multivariate regressions for determination sugar content of intact Gannan navel orange
536 by Vis–NIR diffuse reflectance spectroscopy. Mathematical and Computer Modelling,
537 51(11), 1438-1443.

538 Lopez, A., Arazuri, S., Garcia, I., Mangado, J., & Jaren, C. (2013). A review of the
539 application of near-infrared spectroscopy for the analysis of potatoes. J. Agric. Food
540 Chem., 61, 5413–5424.

Lorber, A., Wangen, L., & Kowalski, B. (1987). A theoretical foundation for the PLS algorithm. Journal of Chemometrics, 1, 19–31.

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.

Matsuo, T. & Ito, S. (1982). 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 46, 683–689.

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

Mowat, A.D. & Poole, P.R. (1997). Non-destructive discrimination of persimmon fruit quality using visible-near infrared reflectance spectrophotometry. Acta Hort. (ISHS) 436, 159–164, http://www.actahort.org/books/436/436_17.htm (accessed January 2017).

Munera, S., Besada, C., Aleixos, A., Talens, P., Salvador, A., Sun,, D-W., Cubero, C., & Blasco, J. (2017b). Non-destructive assessment of the internal quality of intact persimmon using colour and VIS/NIR hyperspectral imaging. LWT - Food Science and Technology, 77C, 241-248.

Munera, S., Besada, C., Blasco, J., Cubero, S., Salvador, A., Talens, P., & Aleixos, N. (2017a). Astringency assessment of persimmon by hyperspectral imaging. Postharvest Biology and Technology, 125, 35-41.

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

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

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

Novillo, P., Salvador, A., Llorca, E., Hernando, I. & Besada, C. (2014). Effect of CO_2 deastringency treatment on flesh disorders induced by mechanical damage in persimmon. Biochemical and microstructural studies. Food Chemistry 145, 453–463.

Noypitak, S., Terdwongworakul, A., Krisanapook, K. & Kasemsumran, S. (2015). Evaluation of astringency and tannin content in ‘Xichu’ persimmons using near infrared spectroscopy. International Journal of Food Properties, 18(5), 1014-1028.

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

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

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

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

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

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

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

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

Theanjumpol, P., Self, G., Rittiron, R., Pankasemsu, T., & Sardsud, V. (2013). Selecting Variables for Near Infrared Spectroscopy (NIRS) Evaluation of Mango Fruit Quality. Journal of Agricultural Science, 5(7), 146-159.

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

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

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

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

Williams, P., & Sobering, D. (1993). Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. J. Near Infrared Spectr., 1, 25-32.

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

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

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

Figure 2. A labelled picture of the spectrometer

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

Figure 4. Normalised X-loading weights of the best PLSR model for the six measurement points (with SNV+1-Der pre-processing for the intact fruit set) for the (a) VNIR and (b) NIR detectors, respectively. Only the weights corresponding to the latent variables that explain 95 % of the Y-variable variance are shown (5 for VNIR and 16 for NIR detectors)

Figure 5. Impressions of tannin content representing the evolution of the astringency distribution and intensity for persimmons after different CO_2 treatments: a) untreated; and b-e) treated with CO_2 for 6, 12, 18 and 24 h, respectively

644 **Table 1**. Statistical values of tannin 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

645

646 **Table 2**. Results of tannin content using the average of the six measurement points with
647 all wavelengths by PLSR, SVM and LS-SVM models for the intact fruit set

Model	Pre-treatment	*LV, γ*	VNIR			*LV, γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	**RAW**	18	0.829	0.100	2.45	36	0.813	0.105	2.34
	SNV	17	0.828	0.101	2.44	35	0.810	0.106	2.32
	SG	19	0.802	0.108	2.28	46	0.758	0.119	2.06
	1-Der	9	0.898	0.077	3.17	28	0.850	0.094	2.61
	2-Der	9	0.885	0.082	2.98	24	0.755	0.120	2.05
	MSC	17	0.828	0.101	2.44	34	0.821	0.103	2.39
	DOSC	1	0.817	0.104	2.37	1	0.704	0.132	1.86
	SNV + 1-Der	10	0.904	0.075	3.26	27	0.861	0.090	2.72
	SNV+ 2-Der	10	0.883	0.083	2.96	22	0.795	0.110	2.23
	SNV+DOSC	1	0.814	0.104	2.35	18	0.814	0.105	2.34
SVM	**RAW**		0.813	0.105	2.34		0.117	0.256	0.96
	SNV		0.863	0.090	2.74		0.010	0.241	1.02
	SG		0.813	0.105	2.34		0.107	0.255	0.96
	1-Der		0.893	0.079	3.09		0.728	0.126	1.94
	2-Der		0.896	0.078	3.14		0.811	0.105	2.33
	MSC		0.861	0.090	2.71		0.016	0.244	1.00
	DOSC		0.835	0.099	2.49		0.731	0.126	1.95
	SNV + 1-Der		0.894	0.079	3.11		0.852	0.093	2.63
	SNV+ 2-Der		0.897	0.078	3.15		0.861	0.090	2.72
	SNV+DOSC		0.834	0.099	2.48		0.899	0.077	3.19
LS-SVM	**RAW**	1.828	0.805	0.107	2.29	4126.52	0.814	0.105	2.35
	SNV	4278.28	0.821	0.102	2.39	59.782	0.870	0.087	2.81
	SG	111.231	0.789	0.111	2.20	4035.02	0.760	0.119	2.07
	1-Der	82.282	0.868	0.088	2.79	1.275	0.805	0.107	2.29
	2-Der	13.288	0.860	0.091	2.71	0.215	0.738	0.124	1.98
	MSC	0.014	0.829	0.100	2.44	80.185	0.862	0.090	2.72
	DOSC	1.35×10^{10}	0.817	0.104	2.37	4.61×10^{13}	0.704	0.132	1.86
	SNV + 1-Der	358.236	0.877	0.085	2.88	89.781	0.866	0.089	2.77
	SNV+ 2-Der	184.810	0.885	0.082	2.99	0.109	0.805	0.107	2.29
	SNV+DOSC	2.10×10^{6}	0.815	0.104	2.35	0.002	0.897	0.078	3.15

648

649

650 **Table 3**. Results of tannin content using the average of the six measurement points with
651 all wavelengths by PLSR, SVM and LS-SVM models for the half cut fruit set

Model	Pre-treatment	*LV, γ*	VNIR			*LV, γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	**RAW**	15	0.761	0.118	2.07	38	0.733	0.125	1.96
	SNV	14	0.741	0.123	1.99	37	0.736	0.125	1.97
	SG	17	0.727	0.127	1.94	59	0.329	0.198	1.24
	1-Der	9	0.856	0.092	2.66	31	0.659	0.142	1.73
	2-Der	9	0.864	0.089	2.74	22	0.583	0.156	1.57
	MSC	14	0.741	0.123	1.99	37	0.729	0.126	1.94
	DOSC	1	0.741	0.123	1.99	1	0.604	0.153	1.61
	SNV + 1-Der	8	0.844	0.096	2.57	30	0.678	0.138	1.78
	SNV+ 2-Der	9	0.861	0.090	2.72	22	0.642	0.145	1.69
	SNV+DOSC	1	0.744	0.123	2.00	7	0.712	0.130	1.88
SVM	**RAW**		0.826	0.101	2.43	0	0.174	0.220	1.11
	SNV		0.813	0.105	2.34	0	0.557	0.161	1.52
	SG		0.792	0.110	2.22	0	0.098	0.230	1.07
	1-Der		0.872	0.087	2.83	0	0.822	0.102	2.40
	2-Der		0.877	0.085	2.88	0	0.841	0.097	2.54
	MSC		0.800	0.108	2.26	0	0.526	0.167	1.47
	DOSC		0.754	0.120	2.04	0	0.629	0.148	1.66
	SNV + 1-Der		0.858	0.091	2.68	0	0.812	0.105	2.33
	SNV+ 2-Der		0.871	0.087	2.82	0	0.853	0.093	2.64
	SNV+DOSC		0.760	0.119	2.06	0	0.826	0.101	2.42
LS-SVM	**RAW**	1.946	0.796	0.109	2.24	1458.98	0.736	0.125	1.97
	SNV	0.004	0.795	0.110	2.23	32.265	0.794	0.110	2.23
	SG	190.193	0.760	0.119	2.07	1334.51	0.655	0.142	1.72
	1-Der	0.011	0.858	0.091	2.69	0.378	0.819	0.103	2.38
	2-Der	32.619	0.870	0.087	2.80	0.049	0.794	0.110	2.23
	MSC	0.003	0.796	0.110	2.24	24.415	0.783	0.113	2.17
	DOSC	3.26×10^{10}	0.741	0.123	1.99	3.58×10^{9}	0.604	0.153	1.61
	SNV + 1-Der	9577.86	0.849	0.094	2.61	0.163	0.795	0.110	2.23
	SNV+ 2-Der	1.15×10^{4}	0.866	0.089	2.76	0.051	0.817	0.104	2.37
	SNV+DOSC	89.830	0.744	0.123	2.00	0.405	0.819	0.103	2.38

652

Table 4. Results of tannin content using the average of the six measurement points with EWs for the models created by PLSR, SVM and LS-SVM for the intact fruits set (*)

Model	Pre-treatment	*EW/LV, EW, EW/γ*	VNIR			*EW/LV, EW, EW/γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	**SNV + 1-Der**	22/22	0.861	0.090	2.72	48/48	0.893	0.079	3.10
	SNV+ 2-Der	26/26	0.891	0.080	3.06	54/54	0.822	0.102	2.40
	SNV+DOSC	1/1	0.871	0.087	2.81	41/41	0.915	0.071	3.46
SVM	**SNV + 1-Der**	22	0.849	0.094	2.61	48	0.761	0.118	2.07
	SNV+ 2-Der	26	0.884	0.082	2.98	54	0.768	0.117	2.10
	SNV+DOSC	1	0.878	0.085	2.89	41	0.895	0.079	3.12
LS-SVM	**SNV + 1-Der**	22/9.06 x 10^4	0.821	0.103	2.39	48/10.309	0.833	0.099	2.48
	SNV+ 2-Der	26/0.982	0.889	0.081	3.04	54/50.492	0.836	0.098	2.50
	SNV+DOSC	1/122.96	0.874	0.086	2.85	41/3.818	0.893	0.079	3.09

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

Table 5. Results of tannin content using the average of the six measurement points with EWs for the models created by PLSR, SVM and LS-SVM for the half cut fruit set(*)

Model	Pre-treatment	*EW/LV, EW, EW/γ*	VNIR			*EW/LV, EW, EW/γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	**SNV + 1-Der**	30/30	0.880	0.084	2.92	28/28	0.834	0.099	2.48
	SNV+ 2-Der	25/25	0.880	0.084	2.92	38/38	0.790	0.111	2.21
	SNV+DOSC	1/1	0.856	0.092	2.67	51/51	0.850	0.094	2.62
SVM	**SNV + 1-Der**	30	0.879	0.084	2.91	28	0.837	0.098	2.51
	SNV+ 2-Der	25	0.894	0.079	3.12	38	0.743	0.123	2.00
	SNV+DOSC	1	0.862	0.090	2.72	51	0.828	0.101	2.44
LS-SVM	**SNV + 1-Der**	30/0.288	0.865	0.089	2.76	28/8.152	0.774	0.115	2.13
	SNV+ 2-Der	25/2.468	0.885	0.082	2.98	38/6.694	0.743	0.123	2.00
	SNV+DOSC	1/97.163	0.857	0.092	2.68	51/0.030	0.825	0.101	2.42

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

662 **Table 6.** Results of tannin content achieved using different combinations of
663 measurement points and pre-processing methods with all wavelengths by PLSR, SVM
664 and LS-SVM models(*)

Points	Model	Pre-reatment	Entire				BEST PRE-TREAT.	Half cut			
			REG.	R^2	RMSE	RPD		REG.	R^2	RMSE	RPD
1-6-2-5	**PLSR**	**1-Der**	VNIR	0.885	0.082	2.98	**SNV+1-Der**	VNIR	0.829	0.100	2.45
	SVM	**SNV+ 1-Der**	VNIR	0.894	0.079	3.11	**2-Der**	VNIR	0.860	0.091	2.70
	LS-SVM	**SNV+1-Der**	VNIR	0.885	0.082	2.99	**2-Der**	VNIR	0.851	0.094	2.62
1-6-3-4	**PLSR**	**SNV+1-Der**	VNIR	0.884	0.083	2.97	**SNV+2-Der**	VNIR	0.863	0.090	2.73
	SVM	**SNV+ 1-Der**	VNIR	0.885	0.082	2.99	**2-Der**	VNIR	0.883	0.083	2.96
	LS-SVM	**SNV+ 2-Der**	VNIR	0.874	0.086	2.85	**2-Der**	VNIR	0.871	0.087	2.82
1-6	**PLSR**	**SNV+ 1-Der**	VNIR	0.815	0.104	2.35	**SNV+1-Der**	VNIR	0.803	0.108	2.28
	SVM	**SNV+ 1-Der**	VNIR	0.857	0.092	2.67	**2-Der**	VNIR	0.848	0.094	2.60
	LS-SVM	**SNV+ 1-Der**	VNIR	0.843	0.096	2.56	**SNV+2-Der**	VNIR	0.842	0.096	2.54
2-5	**PLSR**	**2-Der**	VNIR	0.869	0.088	2.80	**SNV+1-Der**	VNIR	0.786	0.112	2.19
	SVM	**SNV+ 1-Der**	VNIR	0.882	0.083	2.94	**SNV+2-Der**	NIR	0.837	0.098	2.51
	LS-SVM	**1-Der**	VNIR	0.866	0.089	2.77	**1-Der**	NIR	0.814	0.104	2.35
3-4	**PLSR**	**SNV+ 1-Der**	VNIR	0.837	0.098	2.51	**2-Der**	VNIR	0.852	0.093	2.63
	SVM	**SNV+ 2-Der**	VNIR	0.872	0.087	2.82	**SNV+2-Der**	NIR	0.853	0.093	2.64
	LS-SVM	**SNV+ 2-Der**	VNIR	0.863	0.090	2.73	**1-Der**	VNIR	0.843	0.096	2.55

665 * Only the best prediction results for each model are shown, indicating the associated pre-processing
666
667

Table 7. Results of tannin content achieved using the average of the four measurement points (2-5-3-4) with all wavelengths by PLSR, SVM and LS-SVM models for the intact fruit set(*)

Model	Pre-treatment	*LV, γ*	VNIR			*LV, γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	**SNV + 1-Der**	9	0.874	0.086	2.86	27	0.830	0.100	2.46
	SNV+ 2-Der	9	0.889	0.081	3.04	21	0.760	0.119	2.07
	SNV+DOSC	1	0.808	0.106	2.31	15	0.810	0.106	2.32
SVM	**SNV + 1-Der**		0.895	0.079	3.12		0.862	0.090	2.72
	SNV+ 2-Der		0.890	0.080	3.06		0.813	0.105	2.34
	SNV+DOSC		0.824	0.102	2.41		0.857	0.092	2.68
LS-SVM	**SNV + 1-Der**	4.880	0.872	0.087	2.83	0.230	0.851	0.093	2.62
	SNV+ 2-Der	547.70	0.872	0.087	2.83	0.073	0.760	0.119	2.07
	SNV+DOSC	1.04×10^7	0.808	0.106	2.31	0.001	0.858	0.091	2.68

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

Table 8. Results of tannin content achieved using the average of the four measurement points (2-5-3-4) with all wavelengths by PLSR, SVM and LS-SVM models for the half cut fruit set(*)

Model	Pre-treatment	*LV, γ*	VNIR			*LV, γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	**SNV + 1-Der**	8	0.843	0.096	2.55	30	0.627	0.148	1.66
	SNV+ 2-Der	8	0.827	0.101	2.43	19	0.765	0.117	2.09
	SNV+DOSC	1	0.712	0.130	1.89	7	0.630	0.147	1.66
SVM	**SNV + 1-Der**		0.856	0.092	2.66		0.827	0.101	2.43
	SNV+ 2-Der		0.834	0.099	2.49		0.877	0.085	2.88
	SNV+DOSC		0.725	0.127	1.93		0.783	0.113	2.17
LS-SVM	**SNV + 1-Der**	2952	0.861	0.091	2.71	1.876	0.812	0.105	2.33
	SNV+ 2-Der	54.177	0.834	0.099	2.48	0.067	0.839	0.097	2.52
	SNV+DOSC	2.62×10^6	0.713	0.130	1.89	12.150	0.761	0.119	2.07

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

678 **Table 9**. Results of tannin content achieved using the average of the four measurement
679 points (2-5-3-4) by PLSR, SVM and LS-SVM models with EWs selected by SPA for
680 the intact fruit set(*)

Model	Pre-treatment	*EW/LV, EW, EW/γ*	VNIR			*EW/LV, EW, EW/γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	**SNV + 1-Der**	16/16	0.838	0.098	2.51	28/28	0.856	0.092	2.67
	SNV+ 2-Der	30/30	0.854	0.093	2.65	28/28	0.779	0.114	2.15
	SNV+DOSC	1/1	0.860	0.091	2.70	30/30	0.865	0.089	2.76
SVM	**SNV + 1-Der**	16	0.851	0.094	2.62	28	0.759	0.119	2.06
	SNV+ 2-Der	30	0.864	0.089	2.74	28	0.806	0.107	2.30
	SNV+DOSC	1	0.862	0.090	2.73	30	0.857	0.092	2.68
LS-SVM	**SNV + 1-Der**	16/0.317	0.834	0.099	2.49	28/4.435	0.813	0.105	2.34
	SNV+ 2-Der	30/0.144	0.843	0.096	2.56	28/0.823	0.749	0.122	2.02
	SNV+DOSC	1/5.785	0.861	0.090	2.71	30/0.009	0.855	0.092	2.65

681 * Only the best prediction results for each model are shown, indicating the associated pre-processing
682
683 **Table 10**. Results of tannin content achieved using the average of the four measurement
684 points (2-5-3-4) by PLSR, SVM and LS-SVM models with EWs selected by SPA for
685 the half cut fruit set(*)

Model	Pre-treatment	*EW/LV, EW, EW/γ*	VNIR			*EW/LV, EW, EW/γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	**SNV + 1-Der**	23/23	0.865	0.089	2.75	28/28	0.823	0.102	2.41
	SNV+ 2-Der	18/18	0.835	0.098	2.49	17/17	0.798	0.109	2.25
	SNV+DOSC	1/1	0.814	0.104	2.35	57/57	0.805	0.107	2.29
SVM	**SNV + 1-Der**	23	0.859	0.091	2.70	28	0.826	0.101	2.43
	SNV+ 2-Der	18	0.811	0.105	2.33	17	0.818	0.103	2.37
	SNV+DOSC	1	0.823	0.102	2.41	57	0.770	0.116	2.11
LS-SVM	**SNV + 1-Der**	23/0.249	0.860	0.091	2.70	28/1.30 x 105	0.805	0.107	2.29
	SNV+ 2-Der	18/44.110	0.835	0.098	2.49	17/39.054	0.775	0.115	2.13
	SNV+DOSC	1/18.698	0.815	0.104	2.36	57/0.051	0.756	0.120	2.05

686 * Only the best prediction results for each model are shown, indicating the associated pre-processing
687
688

Highlights

- Persimmon astringency can be assessed by chemometrics and spectroscopy tecnology
- VIS-NIR in the range 600-1100 nm and NIR in the range 900-1800 nm have been tested
- Several pre-processing and statistical methods have been tested in intact and half cut fruit
- Easy handling applicability of non-destructive technique in internal quality analysis

Figure3a

Click here to download high resolution image

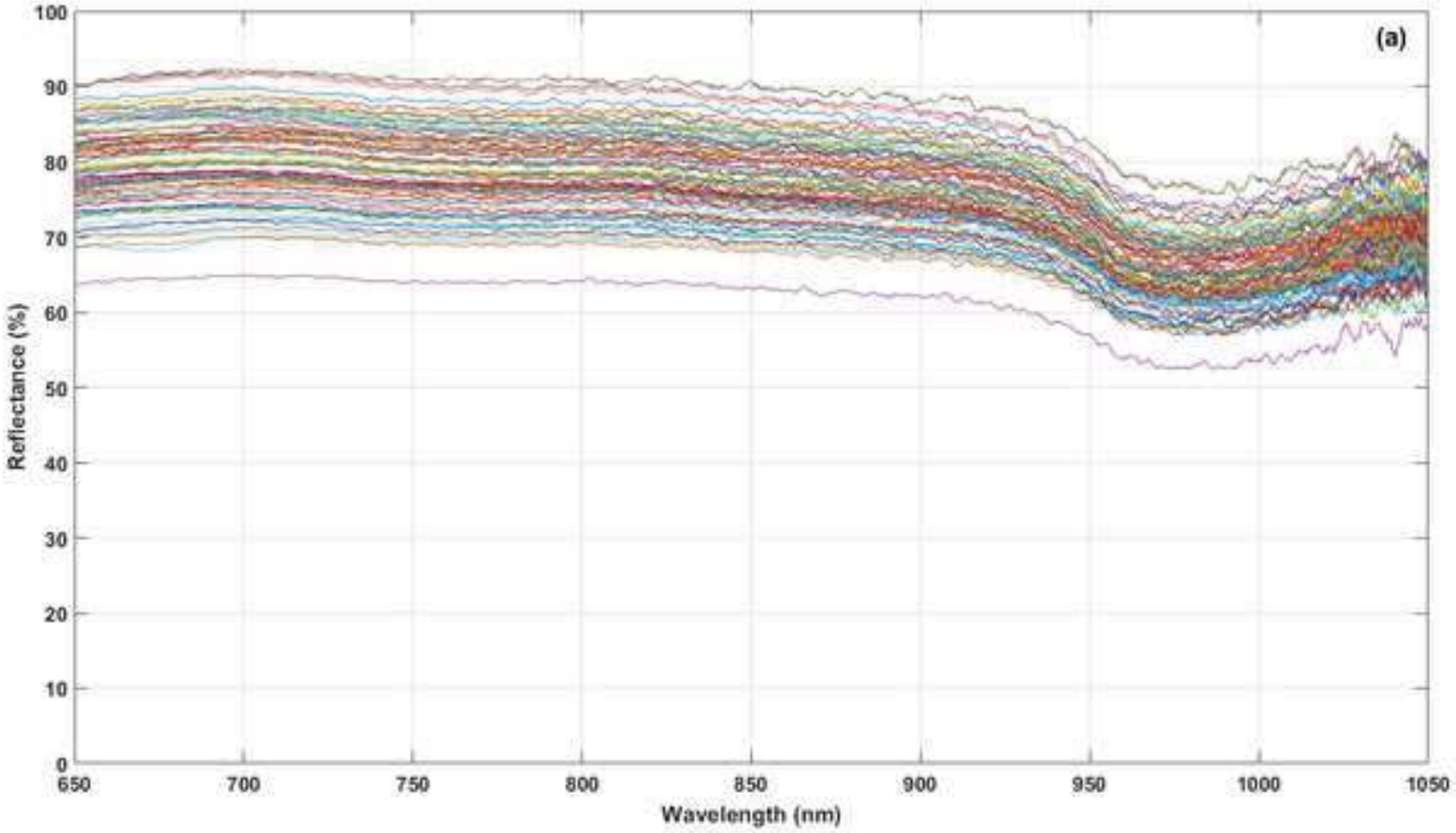

Figure3b

Click here to download high resolution image

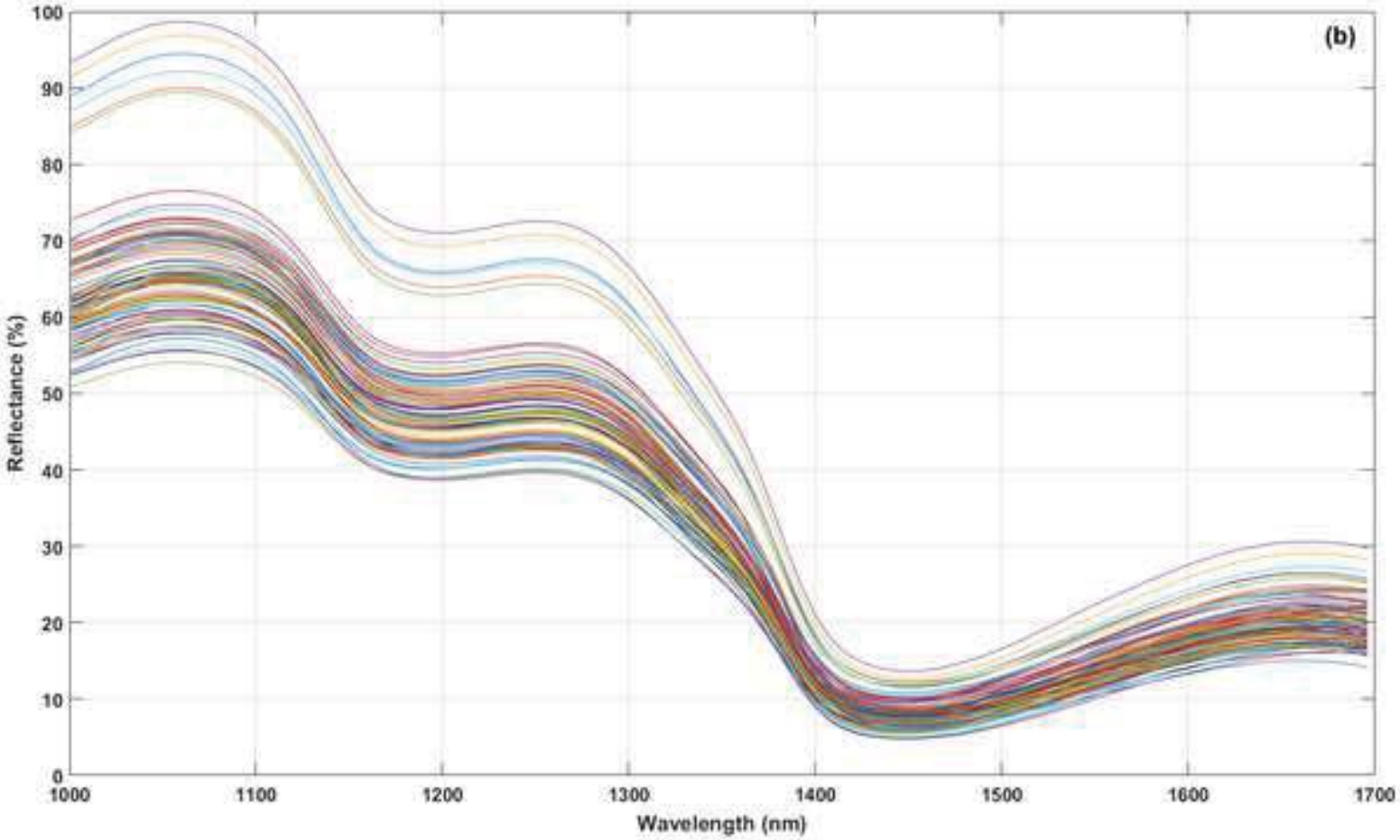

Figure4a

Click here to download high resolution image

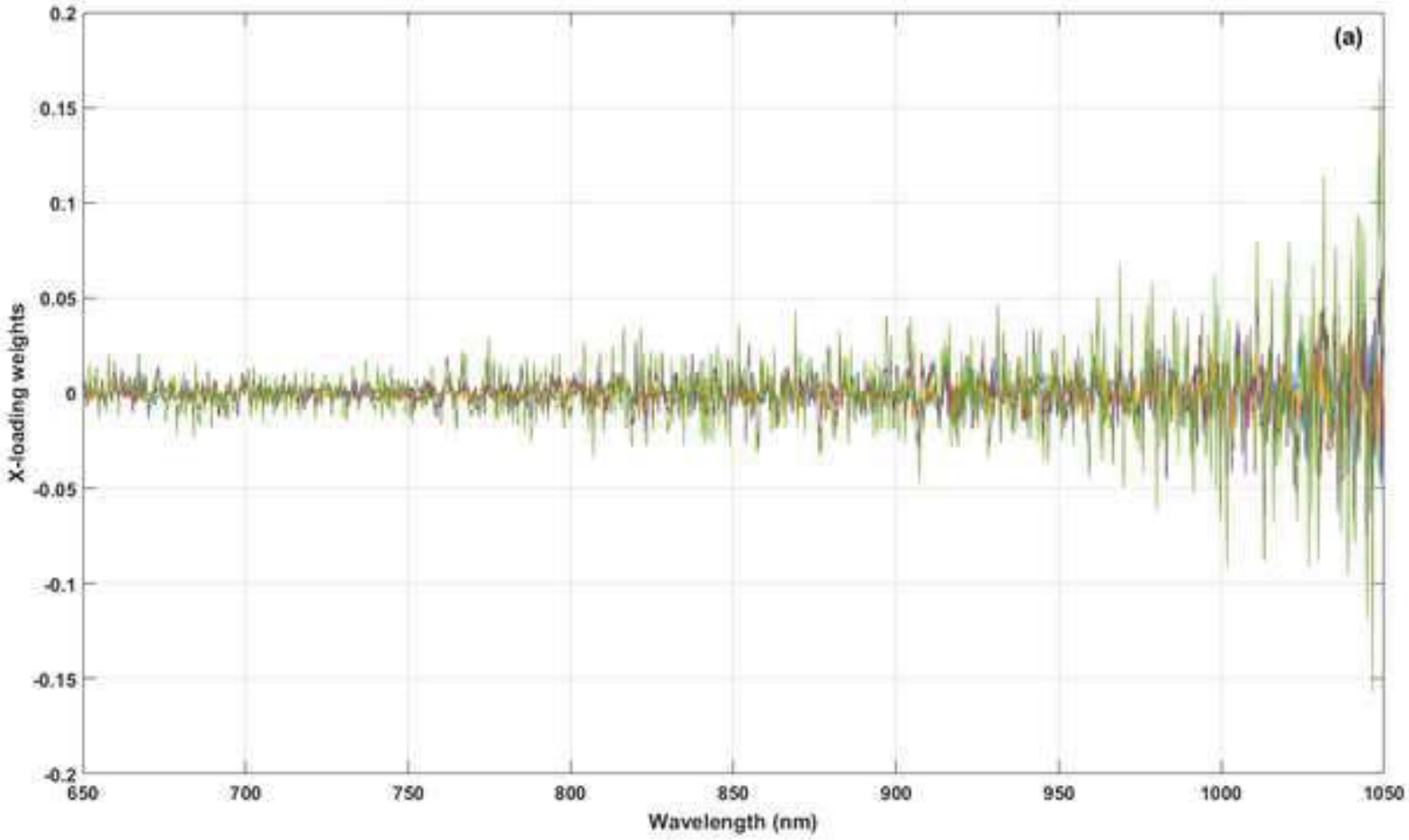

Figure1
Click here to download high resolution image

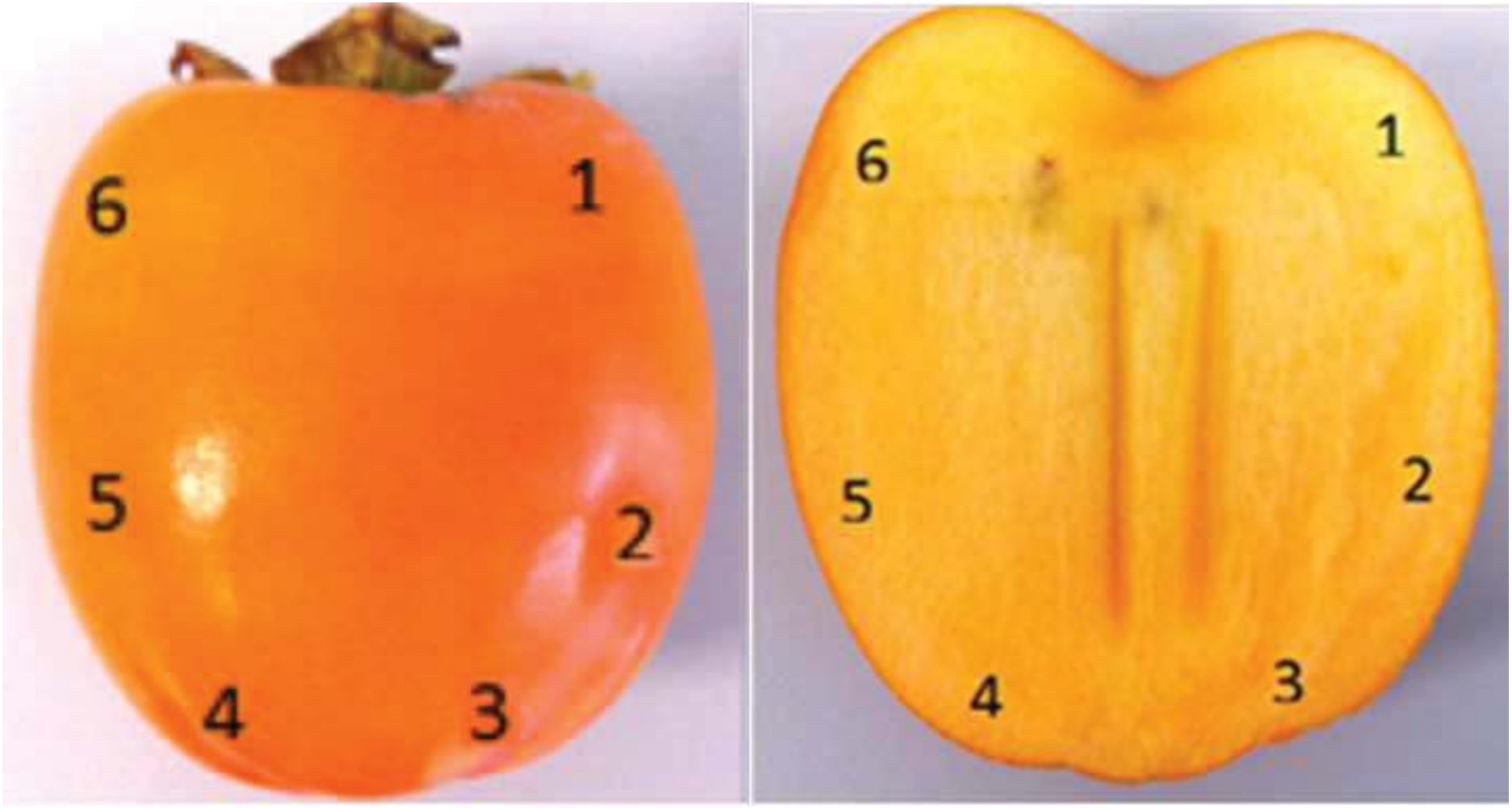

Figure 1. Selected points for the spectroscopic measurements in the: a) intact fruit; and b) the flesh of half cut fruit, respectively

Figure2
Click here to download high resolution image

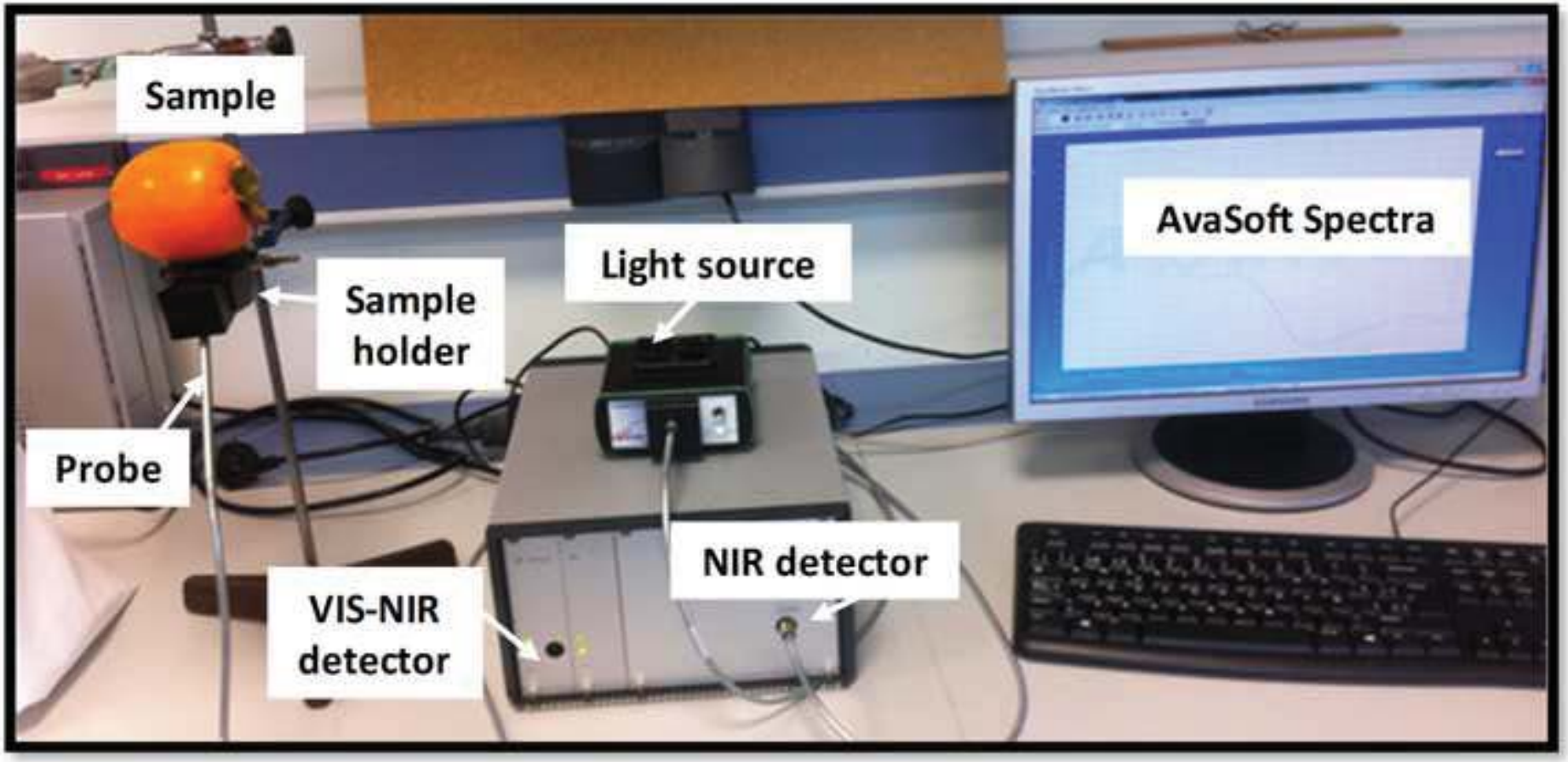

Figure 2. A labelled picture of the VIS/NIR equipment

Figure5
Click here to download high resolution image

a b c d e

Figure 5. 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	.023	.735	.243	.210
Prediction	42	.023	.752	.266	.245

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

MODEL	PRE-TREATMENT	*LV, γ*	VIS-NIR			*LV, γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	***RAW***	18	.829	.100	2.45	36	.813	.105	2.34
	SNV	17	.828	.101	2.44	35	.810	.106	2.32
	SG	19	.802	.108	2.28	46	.758	.119	2.06
	1-Der	9	.898	.077	3.17	28	.850	.094	2.61
	2-Der	9	.885	.082	2.98	24	.755	.120	2.05
	MSC	17	.828	.101	2.44	34	.821	.103	2.39
	DOSC	1	.817	.104	2.37	1	.704	.132	1.86
	SNV + 1-Der	10	.904	.075	3.26	27	.861	.090	2.72
	SNV+ 2-Der	10	.883	.083	2.96	22	.795	.110	2.23
	SNV+DOSC	1	.814	.104	2.35	18	.814	.105	2.34
SVM	***RAW***		.813	.105	2.34		.117	.256	.96
	SNV		.863	.090	2.74		.010	.241	1.02
	SG		.813	.105	2.34		.107	.255	.96
	1-Der		.893	.079	3.09		.728	.126	1.94
	2-Der		.896	.078	3.14		.811	.105	2.33
	MSC		.861	.090	2.71		.016	.244	1.00
	DOSC		.835	.099	2.49		.731	.126	1.95
	SNV + 1-Der		.894	.079	3.11		.852	.093	2.63
	SNV+ 2-Der		.897	.078	3.15		.861	.090	2.72
	SNV+DOSC		.834	.099	2.48		.899	.077	3.19
LS-SVM	***RAW***	1.828	.805	.107	2.29	4126.52	.814	.105	2.35
	SNV	4278.28	.821	.102	2.39	59.782	.870	.087	2.81
	SG	111.231	.789	.111	2.20	4035.02	.760	.119	2.07
	1-Der	82.282	.868	.088	2.79	1.275	.805	.107	2.29
	2-Der	13.288	.860	.091	2.71	.215	.738	.124	1.98
	MSC	.014	.829	.100	2.44	80.185	.862	.090	2.72
	DOSC	1.35×10^{10}	.817	.104	2.37	4.61×10^{13}	.704	.132	1.86
	SNV + 1-Der	358.236	.877	.085	2.88	89.781	.866	.089	2.77
	SNV+ 2-Der	184.810	.885	.082	2.99	.109	.805	.107	2.29
	SNV+DOSC	2.10×10^{6}	.815	.104	2.35	.002	.897	.078	3.15

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

MODEL	PRE-TREATMENT	*LV, γ*	VIS-NIR			*LV, γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	***RAW***	15	.761	.118	2.07	38	.733	.125	1.96
	SNV	14	.741	.123	1.99	37	.736	.125	1.97
	SG	17	.727	.127	1.94	59	.329	.198	1.24
	1-Der	9	.856	.092	2.66	31	.659	.142	1.73
	2-Der	9	.864	.089	2.74	22	.583	.156	1.57
	MSC	14	.741	.123	1.99	37	.729	.126	1.94
	DOSC	1	.741	.123	1.99	1	.604	.153	1.61
	SNV + 1-Der	8	.844	.096	2.57	30	.678	.138	1.78
	SNV+ 2-Der	9	.861	.090	2.72	22	.642	.145	1.69
	SNV+DOSC	1	.744	.123	2.00	7	.712	.130	1.88
SVM	***RAW***		.826	.101	2.43		.174	.220	1.11
	SNV		.813	.105	2.34		.557	.161	1.52
	SG		.792	.110	2.22		.098	.230	1.07
	1-Der		.872	.087	2.83		.822	.102	2.40
	2-Der		.877	.085	2.88		.841	.097	2.54
	MSC		.800	.108	2.26		.526	.167	1.47
	DOSC		.754	.120	2.04		.629	.148	1.66
	SNV + 1-Der		.858	.091	2.68		.812	.105	2.33
	SNV+ 2-Der		.871	.087	2.82		.853	.093	2.64
	SNV+DOSC		.760	.119	2.06		.826	.101	2.42
LS-SVM	***RAW***	1.946	.796	.109	2.24	1458.98	.736	.125	1.97
	SNV	.004	.795	.110	2.23	32.265	.794	.110	2.23
	SG	190.193	.760	.119	2.07	1334.51	.655	.142	1.72
	1-Der	.011	.858	.091	2.69	.378	.819	.103	2.38
	2-Der	32.619	.870	.087	2.80	.049	.794	.110	2.23
	MSC	.003	.796	.110	2.24	24.415	.783	.113	2.17
	DOSC	3.26×10^{10}	.741	.123	1.99	3.58×10^{9}	.604	.153	1.61
	SNV + 1-Der	9577.86	.849	.094	2.61	.163	.795	.110	2.23
	SNV+ 2-Der	1.15×10^{4}	.866	.089	2.76	.051	.817	.104	2.37
	SNV+DOSC	89.830	.744	.123	2.00	.405	.819	.103	2.38

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

MODEL	PRE-TREATMENT	*EW/LV, EW, EW/γ*	VIS-NIR			*EW/LV, EW, EW/γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	***SNV + 1-Der***	22/22	.861	.090	2.72	48/48	.893	.079	3.10
	SNV+ 2-Der	26/26	.891	.080	3.06	54/54	.822	.102	2.40
	SNV+DOSC	1/1	.871	.087	2.81	41/41	.915	.071	3.46
SVM	***SNV + 1-Der***	22	.849	.094	2.61	48	.761	.118	2.07
	SNV+ 2-Der	26	.884	.082	2.98	54	.768	.117	2.10
	SNV+DOSC	1	.878	.085	2.89	41	.895	.079	3.12
LS-SVM	***SNV + 1-Der***	22/9.06 x 10^4	.821	.103	2.39	48/10.309	.833	.099	2.48
	SNV+ 2-Der	26/.982	.889	.081	3.04	54/50.492	.836	.098	2.50
	SNV+DOSC	1/122.96	.874	.086	2.85	41/3.818	.893	.079	3.09

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

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

MODEL	PRE-TREATMENT	*EW/LV, EW, EW/γ*	VIS-NIR			*EW/LV, EW, EW/γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	***SNV + 1-Der***	30/30	.880	.084	2.92	28/28	.834	.099	2.48
	SNV+ 2-Der	25/25	.880	.084	2.92	38/38	.790	.111	2.21
	SNV+DOSC	1/1	.856	.092	2.67	51/51	.850	.094	2.62
SVM	***SNV + 1-Der***	30	.879	.084	2.91	28	.837	.098	2.51
	SNV+ 2-Der	25	.894	.079	3.12	38	.743	.123	2.00
	SNV+DOSC	1	.862	.090	2.72	51	.828	.101	2.44
LS-SVM	***SNV + 1-Der***	30/.288	.865	.089	2.76	28/8.152	.774	.115	2.13
	SNV+ 2-Der	25/2.468	.885	.082	2.98	38/6.694	.743	.123	2.00
	SNV+DOSC	1/97.163	.857	.092	2.68	51/.030	.825	.101	2.42

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

Table 6. Prediction results of tannins content using different combinations of measuring points and pre-processing with all wavelengths by PLSR, SVM and LS-SVM models (*)

POINTS	MODEL	BEST PRE-TREAT.	Entire				BEST PRE-TREAT.	Half cut			
			REG.	R^2	RMSE	RPD		REG.	R^2	RMSE	RPD
1-6-2-5	PLSR	*1-Der*	VIS-NIR	.885	.082	2.98	*SNV + 1-Der*	VIS-NIR	.829	.100	2.45
	SVM	*SNV + 1-Der*	VIS-NIR	.894	.079	3.11	*2-Der*	VIS-NIR	.860	.091	2.70
	LS-SVM	*SNV+1-Der*	VIS-NIR	.885	.082	2.99	*2-Der*	VIS-NIR	.851	.094	2.62
1-6-3-4	PLSR	*SNV+1-Der*	VIS-NIR	.884	.083	2.97	*SNV + 2-Der*	VIS-NIR	.863	.090	2.73
	SVM	*SNV + 1-Der*	VIS-NIR	.885	.082	2.99	*2-Der*	VIS-NIR	.883	.083	2.96
	LS-SVM	*SNV + 2-Der*	VIS-NIR	.874	.086	2.85	*2-Der*	VIS-NIR	.871	.087	2.82
1-6	PLSR	*SNV + 1-Der*	VIS-NIR	.815	.104	2.35	*SNV + 1-Der*	VIS-NIR	.803	.108	2.28
	SVM	*SNV + 1-Der*	VIS-NIR	.857	.092	2.67	*2-Der*	VIS-NIR	.848	.094	2.60
	LS-SVM	*SNV + 1-Der*	VIS-NIR	.843	.096	2.56	*SNV + 2-Der*	VIS-NIR	.842	.096	2.54
2-5	PLSR	*2-Der*	VIS-NIR	.869	.088	2.80	*SNV + 1-Der*	VIS-NIR	.786	.112	2.19
	SVM	*SNV + 1-Der*	VIS-NIR	.882	.083	2.94	*SNV + 2-Der*	NIR	.837	.098	2.51
	LS-SVM	*1-Der*	VIS-NIR	.866	.089	2.77	*1-Der*	NIR	.814	.104	2.35
3-4	PLSR	*SNV + 1-Der*	VIS-NIR	.837	.098	2.51	*2-Der*	VIS-NIR	.852	.093	2.63
	SVM	*SNV + 2-Der*	VIS-NIR	.872	.087	2.82	*SNV + 2-Der*	NIR	.853	.093	2.64
	LS-SVM	*SNV + 2-Der*	VIS-NIR	.863	.090	2.73	*1-Der*	VIS-NIR	.843	.096	2.55

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

Table 7. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with all wavelengths by PLSR, SVM and LS-SVM models for the intact fruit set (*)

MODEL	PRE-TREATMENT	*LV, γ*	VIS-NIR			*LV, γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	***SNV + 1-Der***	9	.874	.086	2.86	27	.830	.100	2.46
	SNV+ 2-Der	9	.889	.081	3.04	21	.760	.119	2.07
	SNV+DOSC	1	.808	.106	2.31	15	.810	.106	2.32
SVM	***SNV + 1-Der***		.895	.079	3.12		.862	.090	2.72
	SNV+ 2-Der		.890	.080	3.06		.813	.105	2.34
	SNV+DOSC		.824	.102	2.41		.857	.092	2.68
LS-SVM	***SNV + 1-Der***	4.880	.872	.087	2.83	.230	.851	.093	2.62
	SNV+ 2-Der	547.70	.872	.087	2.83	.073	.760	.119	2.07
	SNV+DOSC	1.04×10^7	.808	.106	2.31	.001	.858	.091	2.68

* 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 PLSR, SVM and LS-SVM models for the half cut fruit set(*)

MODEL	PRE-TREATMENT	*LV, γ*	VIS-NIR			*LV, γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	***SNV + 1-Der***	8	.843	.096	2.55	30	.627	.148	1.66
	SNV+ 2-Der	8	.827	.101	2.43	19	.765	.117	2.09
	SNV+DOSC	1	.712	.130	1.89	7	.630	.147	1.66
SVM	***SNV + 1-Der***		.856	.092	2.66		.827	.101	2.43
	SNV+ 2-Der		.834	.099	2.49		.877	.085	2.88
	SNV+DOSC		.725	.127	1.93		.783	.113	2.17
LS-SVM	***SNV + 1-Der***	2952	.861	.091	2.71	1.876	.812	.105	2.33
	SNV+ 2-Der	54.177	.834	.099	2.48	.067	.839	.097	2.52
	SNV+DOSC	2.62×10^6	.713	.130	1.89	12.150	.761	.119	2.07

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

Table 9. Prediction results of tannins content using the average of the four measuring points (2-5-3-4) with EWs (selected by SPA) by PLSR, SVM and LS-SVM models for the intact fruit set(*)

MODEL	PRE-TREATMENT	*EW/LV, EW, EW/γ*	VIS-NIR			*EW/LV, EW, EW/γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	***SNV + 1-Der***	16/16	.838	.098	2.51	28/28	.856	.092	2.67
	SNV+ 2-Der	30/30	.854	.093	2.65	28/28	.779	.114	2.15
	SNV+DOSC	1/1	.860	.091	2.70	30/30	.865	.089	2.76
SVM	***SNV + 1-Der***	16	.851	.094	2.62	28	.759	.119	2.06
	SNV+ 2-Der	30	.864	.089	2.74	28	.806	.107	2.30
	SNV+DOSC	1	.862	.090	2.73	30	.857	.092	2.68
LS-SVM	***SNV + 1-Der***	16/.317	.834	.099	2.49	28/4.435	.813	.105	2.34
	SNV+ 2-Der	30/.144	.843	.096	2.56	28/.823	.749	.122	2.02
	SNV+DOSC	1/5.785	.861	.090	2.71	30/.009	.855	.092	2.65

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

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 PLSR, SVM and LS-SVM models for the half cut fruit set(*)

MODEL	PRE-TREATMENT	*EW/LV, EW, EW/γ*	VIS-NIR			*EW/LV, EW, EW/γ*	NIR		
			R^2	RMSE	RPD		R^2	RMSE	RPD
PLSR	***SNV + 1-Der***	23/23	.865	.089	2.75	28/28	.823	.102	2.41
	SNV+ 2-Der	18/18	.835	.098	2.49	17/17	.798	.109	2.25
	SNV+DOSC	1/1	.814	.104	2.35	57/57	.805	.107	2.29
SVM	***SNV + 1-Der***	23	.859	.091	2.70	28	.826	.101	2.43
	SNV+ 2-Der	18	.811	.105	2.33	17	.818	.103	2.37
	SNV+DOSC	1	.823	.102	2.41	57	.770	.116	2.11
LS-SVM	***SNV + 1-Der***	23/.249	.860	.091	2.70	28/1.30 x 105	.805	.107	2.29
	SNV+ 2-Der	18/44.110	.835	.098	2.49	17/39.054	.775	.115	2.13
	SNV+DOSC	1/18.698	.815	.104	2.36	57/.051	.756	.120	2.05

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

Figure4b

Click here to download high resolution image