

Firmness prediction in *Prunus persica* 'Calrico' peaches by Visible/short wave near infrared spectroscopy and acoustic measurements using optimized linear and non-linear chemometric models

Victoria Lafuente^{1,2}, Luis Javier Herrera³, M^a del Mar Pérez⁴, Jesús Val¹, Ángel Ignacio Negueruela²

¹Estación Experimental de Aula Dei (CSIC). Departamento de Nutrición Vegetal. Avda. Montañana, 1005 C.P. 50059, Zaragoza

²Departamento de Física Aplicada. Facultad de Veterinaria. Universidad de Zaragoza. Calle Miguel Servet 177, C.P. 50013, Zaragoza

³Departamento de Arquitectura y Tecnología de los Computadores. C.I.T.I.C. Universidad de Granada, C.P. 18071, Granada

⁴Departamento de Óptica. Universidad de Granada. C.P. 18002, Granada

mvlafuente@ipe.csic.es

Abstract

In this work, near infrared spectroscopy (NIR) and an acoustic measure (AWETA) (two non-destructive methods) were applied in *Prunus persica* fruit 'Calrico' (n=260) to predict Magness-Taylor (MT) firmness. Separate and combined use of these measures was evaluated and compared using PLS and LS-SVM regression methods. Also, a Mutual Information (MI)-based variable selection method, seeking to find the most significant variables to produce optimal accuracy of the regression models, was applied to a joint set of variables (NIR wavelengths and AWETA measure). The newly proposed combined NIR-AWETA model gave good values of the determination coefficient (R^2) for PLS and LS-SVM methods (0.77 and 0.78, respectively), improving the reliability of MT firmness prediction in comparison with separate NIR and AWETA predictions. The three variables selected by the variable selection method (AWETA measure plus NIR wavelengths 675 and 697 nm) achieved R^2 values 0.76 and 0.77, PLS and LS-SVM. These results indicated that the proposed MI-based variable selection algorithm was a powerful tool for the selection of the most relevant variables.

Keywords: NIR, acoustic system, firmness, peaches

INTRODUCTION

From the time the fruit of peach (*Prunus persica*) is harvested, it is separated from its source of nutrients but the fruit tissues continue breathing and are physiologically active. Ripening is linked to complex physical and chemical changes, such as softening, increased concentration soluble sugars, flavour and colour changes. These processes are important because they influence changes that occur during storage, transportation, and marketing and to some extent affect its nutritional value and sensorial characteristics¹. In addition, these metabolic processes are influenced by external factors such as seasonal and genetic factors, e.g. crop variety or situation. Harvest date not only influences product quality but also boosts production while minimizing agricultural costs².

Firmness is one of the parameters that correlates best with fruit ripeness and is ultimately very useful in determining the optimal harvest date^{3,4}. Firmness can be defined as the resistance by the fruit pulp⁵ to penetration force. Firmness has long been determined by the destructive 'Magness-Taylor' method (MT). The MT test involves attaching a dynamometer to a cylindrical rod, which is inserted 8 mm into the fruit pulp after part of the skin is removed. Although this test is inexpensive and fast, it is of limited usefulness because it destroys the fruit. Also, it renders uncertain values and is highly variable, depending on the force exerted by the operator and fruit spot from where the measurements are taken.

Since 2003, more practical and reproducible non-destructive methods have been developed for estimating fruit firmness with the aim of replacing destructive ones. One such method involves the use of acoustic signals caused by vibrations or mechanical impacts to the fruit. The resonant frequency that is

emitted by an object depends directly on the geometry, mass, and modulus of elasticity of the material being analysed⁶. Acoustical tests performed on the sample are stimulated by a low-intensity impact, producing a vibratory response within an audible range (20-20000 Hertz). The response is recorded with a microphone and the signal in time is processed using Fast Fourier Transform to produce the corresponding signal in the frequency, which produces an acoustic firmness index. Many studies are available on the use of acoustic methods in determining the firmness in various types of fruit and vegetables. However, the results show a high variability in the correlation between the two variables, depending on the cultivar and variety of the analyte (e.g. Apple variety 'Delicious' $R^2 = 0.25$, apple variety 'Golden Delicious' $R^2 = 0.72$ ⁷; avocado $R^2 = 0.43$ ⁸).

Another emerging non-destructive technique currently used for firmness prediction is Near-Infrared Spectroscopy (NIRS). Many researchers are developing methods based on this technique to predict firmness values for different fruits (⁹ apples, ¹⁰ nectarines, ¹¹ peaches, ¹² avocados). In the NIRS technique, a light beam strikes the fruit, penetrating a few millimetres. Part of this radiation is absorbed by the sample and the rest is reflected towards a detector, which converts this light into a spectral characteristic for each sample. Depending on their physical and chemical composition, different compounds such as sugars, water, and pigments absorb part of the radiation in the visible and infrared region.

Signals recorded in these non-destructive methods are related to the desired estimation parameter from which calibration equations are formulated. The resulting firmness MT value is considered to be reference. Despite the variability, it is an acceptable reference in fieldwork. Linear model connections are some of the more commonly used regression methods, such as Multiple Linear Regression' (MLR)¹³, 'Principal Components Regression' (PCR)¹⁴, Partial Least Squares (PLS)¹⁵. Nonlinear methods such as 'Least Squares Support Vector Machines' (LS-SVM)¹⁶ and 'Artificial Neural Networks' (ANN)¹⁷ are used also.

The use of variable selection methods is of particular relevance in these types of physicochemical property-modelling problems related to fruit ripeness. These methods aim to determine the actual contribution of the factors (wavelengths, acoustic frequencies) considered in estimating the objective parameter. These techniques enable the development of predictive models by eliminating spurious factors that complicate the model and that may affect output. In the case of NIRS, different authors have used the selection of variables to improve their results, such as ¹⁶ ($R = 0.97$, RMSEC= 0.2 ° Brix) and ¹⁸ (RMSEP= 0.02 mg/l). These approaches include methods such as Independent Component Analysis (ICA)¹⁶, Successive Projections Algorithm (SPA)¹⁸ and Uninformative Variable Elimination (UVE)¹⁹. In addition, methods based on Mutual Information (MI) are of particular relevance^{20, 21}. These methods use MI as a measure of nonlinear correlation among intervening factors and the variable to be estimated. They have the advantage of not altering the original features considered and directly detect the factors.

The main objective of this work was to investigate peach firmness prediction from the variety 'Calrico' using information gained through non-destructive methods Vis / NIR and acoustic signalling (AWETA). A Mutual-Information-based variable selection method was developed to identify the relevance of both sources of information and select the most prominent factors to make the prediction and build the most accurate and reproducible models.

MATERIAL AND METHODS

Samples

This study was performed during 2010 and 2011, in an experimental peach (*Prunus persica*) orchard, called AFRUCAS, located in the area of "Badilla del Corral de Dolader", belonging to Caspe, (Zaragoza, Spain). The experiment was performed on *peach variety 'Calrico'* over a hybrid pattern. This variety is of particular interest in the area of Bajo Aragón, which has applied for the certificate of origin 'Calanda'. The trees were planted in 2000, in a pattern of 5x3 meters. Fertigation was applied by local drip (25,000 L/ha/h). Samples (260 fruits) were harvested in 2010 (150), 50 samples were harvested weekly for three weeks and in 2011 (110), 22 peaches were tested weekly for five weeks. In this way, maturity ranged from hard green (unripe) to soft mature (overripe) stages. The samples were harvested early in the

morning, transported at 20°C to the lab for analysis (same afternoon). Non-destructive determinations (AWETA and NIRS) were made prior to destructive MT firmness determination.

Vis/NIR spectroscopy

Spectra from intact peaches were measured with a reflectance modular equipment Multispec instrument (AG Tec5, AM Frankfurt, Germany) equipped with a spectrometer SC - NEM I (Zeiss, Jena, Germany) (range: 400-1060 nm, $\Delta\lambda = 1\text{nm}$), a halogen light source (12 V and 20 W), and one fibre optic to conduct the light from source on the sample and another fibre to carry the light reflected by the samples to the detector. The illumination-measurement geometry was set up at 45°/0° (Figure 1)

The measured spectra were the average of 50 spectra (integration time 400ms). Two spectral measurements were made per fruit at two fixed opposed positions on the equator, using the mean of the two spectra for the calibration processes.

Acoustic measurements

Acoustic firmness measurements were made by means of a commercial desktop acoustic-firmness sensor (model AFS, AWETA, The Netherlands). The sensor recorded the weight and resonant frequencies of the acoustic vibration generated by gently tapping the fruit on the equatorial area, from which an acoustic firmness index was provided by AWETA equipment^{22, 23, 24}

$$FI = f_R^2 m^{2/3}$$

In this index, the mass (m, Kg) and the resonant frequency of maximum amplitude are related (f_R , Hz)

Acoustic measurements were made three times per fruit and the average value was calculated.

Magness-Taylor test

The Magness-Taylor firmness test was performed using a hand-held penetrometer Fruit Pressure Tester FT32 (Istituto per la Valorizzazione dei Prodotti Agricoli, Italy) with a probe 8 mm in diameter. The fruit skin was removed with a blade at two positions around the equator and firmness measured (Kg). The firmness was tested in the same area where NIR reflectance spectra were measured.

Data analysis

Chemometric data analysis was performed using the Matlab R2009a software with own programming. Different pre-treatment combinations were evaluated for calibration models such as MSC (Multiplicative Scatter Correction), SNV (Standard Normal Variate) and derivate in first and second order. Finally the best results were found using normalization with the mean 0 and variance 1 of the values for each wavelength and acoustic measurement. This type of normalization is applied individually to each variable considered, whether it be spectral or otherwise; in this way it is independent of whether or not fusion is variable^{20, 25}

Selection of samples for calibration and validation purposes is one of the critical steps when developing a calibration model. To detect the spectra of outliers, a principal components analysis (PCA) was performed. The samples detected as outliers were not included in the calibration groups. Three different random subdivisions on calibration and validation data were performed, using a ratio 2:1 for calibration and validation, respectively¹⁰

Partial Least Square analysis (PLS)

PLS is a regression method often used to construct prediction models for reference parameters established by destructive methods. A high number of variables which may have high collinearity are used. PLS, which allows the Vis/NIR spectra to correlate with physico-chemical parameters of the fruit, has been used in a considerable number of studies related to the prediction of indices in fruits and vegetables^{26, 27, 28}.

In the present study, calibration models were developed to predict firmness in intact peaches. L-Fold Cross-validation (L=10) was used to avoid overfitting in the development of the calibration models^{3,10,29, 30}.

Least squares support vector machine (LS-SVM)

Least squares support vector machine (LS-SVM) is a regression model that has been used in recent years to predict parameters related to fruit ripening and other chemical and physical properties. It has the ability to deal with linear and nonlinear multivariate modelling and resolving these problems in a relatively quick way³¹. Previous research has demonstrated the potential of this non-linear regression model for several quantitative applications in agro-food products^{32, 11, 33}. In the LS-SVM model development, the radial basis function (RBF) kernel was used. Grid-search and cross-validation (LS-SVMLab,²⁵) were used to achieve the optimal combination of gam (γ) and sig² (δ) hyper-parameters of the model.

In both regression methods, the statistics used to select the most accurate models were: coefficient of determination for cross validation (R_{cv}^2) and standard error of cross validation (SECV)³⁴. Other statistics used were the coefficient of variation (%CV), defined as the percentage ratio of the SECV to the mean value of the reference data for the calibration set, and the RPD (Residual Predictive Deviation) calculated as the ratio of the standard deviation (SD) of the reference data to the SECV³⁴. These latter two statistics facilitate the comparison of results found with sets of different means³⁴.

Predictive models formulated were subsequently subjected to internal validation. The effect of different settings on the performance of the model was evaluated by comparing root mean square error of prediction (RMSEP) determination coefficient in validation (R^2), bias and standard error of prediction (SEP).

Mutual Information Backwards Variable selection

Mutual information (MI) is a non-linear correlation measurement from the Information Theory³⁵. For two sets of continuous variables, X and Y, it can be calculated by:

$$I(X, Y) = \int \frac{\mu_{X,Y}(x,y) \log(\mu_{X,Y}(x,y))}{\mu_X(x)\mu_Y(y)} dx dy$$

where $\mu_{X,Y}(x,y)$ is the joint probability density function (PDF) of X and Y, and $\mu_X(x)$ is the marginal density function of the set of variables X. Its advantage in comparison with other correlation criteria is that it is able to identify non-linear relations among the variables involved.

Several attempts have been reported in recent literature for designing algorithms to identify the most relevant factors (wavelengths and/or other factors) for predicting chemical properties, many of those being based on mutual information^{20, 21, 36}. Variable selection aims at identifying irrelevant and redundant variables for their rejection. Identification of redundant variables is critical in spectrometric problems, as nearby wavelengths provide usually similar information. The reduction in variables needed to predict any magnitude is essential to reduce experimental and evaluation costs, but also for increasing the so-called generalization capability of the models, i.e. prediction capability on unseen data³⁷.

The method proposed in this work is an approach first published³⁸ and adapted for continuous variables³⁷, which is based on the *Markov blanket* concept. Given a set of input variables X and an output

variable Y , a set of variables M_i in X is said to be a *Markov blanket* for a variable x_i in X with respect to Y , if $I(\{M_i \cup x_i\}, Y) = I(M_i, Y)$, provided that M_i itself has all the information that x_i has about Y . A Markov blanket is thus a group of variables that subsumes the mutual information content in a certain variable, in practice (and for our purposes) with respect to the objective variable. The algorithm consists of a backwards variable selection method which starts with the complete set of variables, and iteratively discards those which are detected to have a Markov Blanket in the remaining set X_G of variables—that is, those whose information with respect to Y is already present in the remaining set X_G of variables.

The algorithm states the following steps:

1. Calculate the MI between each pair of input variables $I(x_i, x_j)$
2. Starting from the complete set of input variables $X_G = X$, iterate:
 - a) For each variable x_i , let the candidate Markov blanket M_i be the set of p variables in X_G for which $I(x_i, x_j)$ is highest.
 - b) Compute for each x_i

$$Loss_i = I(\{M_i \cup x_i\}, Y) - I(M_i, Y)$$

- c) Choose the x_i for which $Loss_i$ is lowest and eliminate x_i from X_G .
3. Continue with step 2 until no variables remain.

In this way, the relevance of variables (in reverse order) is ranked. Under this operation, it should be noted that variables that have low influence with respect to the output variable (irrelevant variables) will be soon discarded, as the $Loss_i$ value should tend to 0. Similarly, redundant variables are iteratively discarded at earlier stages. Relevant variables with low redundancy are the last to be “chosen”. Further discussion about efficiency, character and operation of the algorithm can be found ³⁹.

The Mutual Information estimator used in this work is the k-nearest neighbours one ⁴⁰, as it has been shown in the literature to have a more robust behaviour with respect to other alternatives when working with groups of variables. The p parameter of the algorithm (in step 2.a of the algorithm) will take the value $p = 1$, as recommended ³⁹.

RESULTS AND DISCUSSION

Data Description

Table 1 shows the number of calibrated and validated samples. Mean values, standard deviation, range, and coefficient of variation for three subdivisions of the initial set of samples from the MT firmness parameter are also shown for all three groups. It is observed that all three calibration groups present similar mean values, range, and standard deviation. The range of validation of the subgroups is included in those of calibration, which demonstrates that group selection has been made correctly. The values of standard deviation and coefficient of variation indicate high variability in each group.

MT firmness prediction through acoustic measurement

The correlation between the data of MT firmness and acoustic signal was studied. Regression coefficients of the three study groups were between $R^2 = 0.52$ and 0.55 . Internal validation was performed in which R^2 values ranged from 0.45 to 0.58 , the values of $RMSEP = 1.04-1.16$ and $bias = -0.03-0.02$. In similar studies with peaches, values found for the correlation of both parameters were slightly lower ($R^2 = 0.38$) ⁴¹ and $R^2 = 0.45$ ⁴². The results show that, despite having higher R^2 values than previously reported, the acoustic signal had correlations with MT firmness which were relatively low in comparison with other non-destructive methods, such as spectroscopy Vis / NIR ⁴³.

MT firmness prediction through Vis/NIR spectroscopy

Figure 2 shows the average spectra Vis/NIR of peaches at different ripeness stages. The main component of these spectra is the peak chlorophyll absorption (680 nm). Chlorophyll content decreases as fruit ripens. The absorption peak decline was observed until the fruit reached a ripeness point at which the peak disappeared. Changes in the spectral region below 700 nm were attributed to changes in fruit colour, due to pigment absorption, disappearance of chlorophyll and synthesis of coloured pigments, mainly carotenoids and anthocyanins.

Calibration models for MT firmness prediction of peach based on Spectra Vis/NIR were developed with both linear regression (PLS) and nonlinear regression (LS-SVM). Table 2 shows the statistical parameters determined from the analysis through both methods of regression and for the three calibration groups. The R_{cv}^2 values for PLS and LS-SVM had the same value (0.7); a good correlation between spectra and firmness. The nonlinear model LS-SVM presented lower SECV values and higher RPD values for the three calibration groups, signifying that the nonlinear regression method provided a better correlation between MT firmness and the absorption spectrum.

Table 3 shows the statistics of internal validations performed for the three sample groups. R^2 values are similar for both regression methods; however, RMSEP values and %CV are lower for the PLS model. The results imply the possibility of a good MT firmness prediction through the use of NIR spectroscopy, as the findings were consistent with those of other authors in similar works on peaches⁴⁴ ($R_{cv}^2 = 0.74$ and RMSEP = 5.42N) using linear and nonlinear regression models and reporting similar results with both methods²³ in which linear regression methods were used to formulate models, giving $R_c^2 = 0.83$ and SEC = 1.70 kg/cm² as a result.

Regression coefficients from the PLS model for one of the calibration sets are shown in Figure 3. The more relevant wavelengths in the absorption spectrum of peach are identified by higher values, in absolute value, of the regression coefficients. Specifically, it is observed that the bands 470-550 nm and 650-690 nm are of great importance. Bands in 400 and 1060 nm were not considered because they contributed excessive noise and did not improve model. The higher regression coefficient values in the range of 600-700 nm are consistent with previous studies^{9,23,4,11} that consider the visible region to be directly related to fruit ripeness. These absorption bands are related to the pigments that provide the fruit with colour.

MT firmness prediction through the combining of the acoustic signal and spectra Vis/NIR

The acoustic index was combined with each of the spectral vector samples Vis / NIR, giving extended input data (194x661). With the same objective, acoustic measurement was added to the Vis / NIR spectrum. Similar strategies carried out²⁴ with mango ($R_{cv}^2 = 0.61$, RMSECV = 3.20 Nmm) and⁹ with apple ($R_{cv}^2 = 0.80$ to 0.75, SECV = 10.32-11.28 N/cm²) gave favourable results with the combination of variables in all cases.

The results demonstrated the suitability of a new method involving the pooling of heterogeneous information gathered using several methods to construct more effective calibration models for predicting physico-chemical parameters of fruits and vegetables. Therefore, fusion of the firmness index and Vis/NIR spectra are assumed to give more robust predictions for fruit-flesh firmness. Calibration equations were formulated from the proposed regression methods (Table 2) following the same methodology as in the previous cases. R_{cv}^2 values were found to be greater for both methods (LS-SVM, 0.8-0.9 and 0.7 PLS). SECV values were lower for PLS and LS-SVM. RPD values were above 2 in each of the three groups.

Internal validation was also performed (Table 3). RMSEP and coefficients of variation were lower and regression coefficients were higher.

Variable selection

To reduce the number of variables (wavelength and acoustic response) and analyse the possibility of improving the generalization ability of the models constructed, we applied the variable selection method backwards; based on Mutual Information (MI)³⁹. Many researchers use selection methods to detect the most relevant variables for which simpler models and higher outputs result⁴⁵, making use of the correlation coefficient and¹⁸ with the method called Successive Projection Algorithm (SPA). Figure 4 shows the output of the models built through two methods of regression on the number of variables. There is an increase in R^2 with respect to the use of two to three variables. In the third variable, the increase in factors does not imply an immediate output increase.

The regression coefficient calculated by using three variables was 0.77 and 0.73 for LS-SVM and PLS respectively. This pattern is observed in all three groups of calibration. LS-SVM models responded better to the creation of models with fewer variables, as shown in similar works^{46, 16, 3}. This may have occurred because in the nonlinear method, both linear and non-linear are taken, whereas linear models (PLS) work only with a linear relationship between spectral and physicochemical data. Table 3 shows the results of validation of the model constructed with three factors. In both cases for LS-SVM and PLS, the R^2 and RMSEP values were improved and %CV was similar. The main advantages of variable selection reduced collinearity, redundancy, and noise. The use of the spectrum involved building a model with only three variables that produced better results.

From the three variables selected, two correspond to wavelengths of the spectrum Vis/ NIR and the third is the firmness index formulated from the acoustic signal. The two selected wavelengths in the three groups proved similar (679 and 696 nm, 675 and 697 nm, 674 and 693 nm). This confirms the importance of the chlorophyll absorption region to determine MT firmness (see figure 4). Similar results have been reported for apples^{47,48,23}. In this way, to calculate the state of ripeness with much simpler instrumentation that requires only two simple optic devices in the visible region, with a filter and a detector for each wavelength selected. The diffraction grating and detector array are eliminated. AWETA equipment becomes necessary.

CONCLUSIONS

This work, using the peach variety 'Calrico', develops a non-destructive method, increasingly demanded by horticultural and agricultural production centres, for MT firmness. The peach variety treated is of particular interest in the area of Bajo Aragón in Spain, which seeks its certificate of origin 'Calanda'. The proposed method is based on the pooling of non-destructive heterogeneous information, specifically a measure of acoustic pulse and spectral data Vis/NIR.

Initially, the results show that the combination of two non-destructive methods AWETA and Vis/NIR spectroscopy improve the reliability of MT firmness prediction with respect to a single source of information used alone. With this new approach, values R^2 and RMSEP improved in all cases. Furthermore, it was found that the Vis/NIR spectrum provided greater reliability to estimate MT firmness than did the AWETA measure.

Finally, the variable selection method based on mutual information used succeeded in finding three optimal parameters for model calibration from both information sources: the acoustic measurement and two wavelengths and for each the corresponding visible chlorophyll absorption region. While using three variables LS-SVM, maintenance was observed (and even slightly increased) in the output of the calibration models (in comparison with the PLS models). Furthermore a considerable simplification of the constructed prediction models was obtained. The equation formulated uses only 3 of the 661 available input variables, providing a simpler methodology and being more affordable in the field of agriculture.

Lastly, it should be noted that this proposed methodology could be applied to any species and variety of fruit and vegetable, as expected in the future.

ACKNOWLEDGEMENTS

We are indebted to volunteers in this study for their outstanding commitment and cooperation. This work has been partially supported by:

AGL2009-08501/AGR project from National Program of Research Projects.

GENIL-PYR-2014-12 project from the GENIL Program of the CEI BioTic GRANADA

REFERENCES

- ¹ Kader, AA, Flavor quality of fruits and vegetables. *J. Sci. Food Agric.* **88**: 1863-1868 (2008)
- ² Thompson JF, Sistemas de cosecha. In *Tecnología Postcosecha de Productos Hortofrutícolas*. 3ª edición. Kader, A. (Eds.). University of California, Division of Agriculture and Natural Resources, Oakland, California, USA 73-76 (2007)
- ³ Li J, Huang W, Zhao C, Zhang B, A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. *J. Food Eng.* **16**:324-332 (2013)
- ⁴ Ruiz-Altisent M, Lleo L, Riquelme F, Instrumental quality assesment of peaches: Fusion of optical and mechanical parameters. *J. Food Eng.* **74**:490-499 (2006)
- ⁵ Mohsenin N N, Physical properties of plant and animal materials: structure, physical, characteristics and mechanical properties. New York: Gordon and Breach Science Publishers **664** (1986)
- ⁶ Studman C J, Computers and electronics in postharvest technology: a review. *Comp. Elec. Agricul* **30**: 109-124 (2001)
- ⁷ Abbott JA, Massie DR, Upchurch BL, Hruschka WR, Nondestructive sonic firmness measurement of apples. *Transactions of the ASAE* **38**:1461-1466 (1995)
- ⁸ Galili N, Shmulevich I, Benichou N, Acoustic testing of avocado for fruit ripeness evaluation. *Trans. ASAE* **41**:399-407 (1998)
- ⁹ Zude M, Herold B, Roger J M, Bellon-Maurel V, Landahl S, Non-destructive tests on the prediction of apple fruit flesh firmness and soluble solids content on tree and in shelf life. *J. Food Eng.* **77**: 254-260 (2006)
- ¹⁰ Sánchez M., De la Haba MJ, Guerrero JE, Garrido-Varo A, Pérez-Marín D, Testing of a local approach for the prediction of quality parameters in intact nectarines using a portable NIRS instrument. *Postharvest Biol. Technol.* **60**:130-135 (2011)
- ¹¹ Shao Y, Bao Y, He Y, Visible/Near-Infrared Spectra for Linear and Nonlinear Calibrations: A case to predict Soluble Solid Contents and pH Values in Peach. *Food Bioprocess Technol.* **4**: 1376-1383 (2011)

- ¹² Clark CJ, McGlone VA, Requejo C, White A, Woolf AB, Dry matter determination in 'Hass' avocado by NIR spectroscopy. *Postharvest Biol. Technol* **29**: 300-307. (2003)
- ¹³ Clark CJ, McGlone VA, Jordan RB, Detection of brownheart in 'Braeburn' apple by transmission NIR spectroscopy. *Postharvest Biol. Technol* **28**: 87-96 (2002)
- ¹⁴ Hernández-Gómez A, He Y, García-Pereira A, Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *J. Food Eng.* **77**: 313-319 (2006)
- ¹⁵ Peirs A, Tirry J, Verlinden B, Darius P, Nicolai B, Effect of biological variability on the robustness of NIR models for soluble solids content of apples. *Postharvest Biol. Technol* **28**: 269-280 (2003)
- ¹⁶ Liu F, Jiang Y, He Y, Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: A case study to determinate soluble solids content of beer. *Anal. Chim. Acta.* **635**: 45-52 (2009)
- ¹⁷ Kim J, Mowat A, Poole P, Kasabov N, Linear and non-linear pattern recognition models for classification of fruit from visible-near infrared spectra. *Chemom. Intell. Lab. Syst.* **51**: 201-216 (2000)
- ¹⁸ Araújo M, Saldanha T, Galvão R, Yoneyama T, Chame H, Visani V, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.* **57**: 65-73 (2001)
- ¹⁹ Xiaobo Z, Jiewen Z, Povey M, Holmes M, Hanpin M, Variables selection methods in near infrared spectroscopy. *Anal. Chim. Acta.* **667**: 14-32 (2010)
- ²⁰ Rossi F, Lendasse A, François D, Wertz V, Verleysen M, Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. *Chemom. Intell. Lab. Syst.* **80**: 215-226 (2006)
- ²¹ Pent H, Long F, Ding C, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *Trans. Pattern Analysis and Mach Intell* **27**:1226-1238 (2005)
- ²² De Ketelaere B, Howarth S, Crezee L, Lammertyn J, Viaene K, Bulens I, De Baerdemaeker J, Postharvest firmness changes as measured by acoustic and low-mass impact devices: a comparison of techniques. *Postharvest Biol. Technol* **41**:275-284 (2006)
- ²³ Zude-Sasse M, Truppel I, Herold B, An approach to non-destructive apple fruit chlorophyll determination. *Postharvest Biol. Technol* **25**:123-133 (2002)
- ²⁴ Valente M, Leardi R, Self G, Luciano G, Pain JP, Multivariate calibration of mango firmness using vis/NIR spectroscopy and acoustic impulse method. *J. Food Eng.* **94**: 7-13 (2009)
- ²⁵ Suykens JAK, Van Gestel T, De Brabanter J, De Moor J, Vanderwalle B, *Least Squ Supp. Vector Mach*, World Scientific, Singapore (2002)
- ²⁶ Nicolai BM, Theron KI, Lammertyn J, Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple. *Chemom. Intell. Lab. Syst.* **85**:243-252 (2006)
- ²⁷ Lu R, Predicting firmness and sugar content of sweet cherries using near infrared difusse reflectance spectroscopy. *Trans ASAE* **44**:1265-1271 (2001)

- ²⁸ Pérez-Marín D, Sánchez MT, Cano G, Garrido A, Prediction of texture in green asparagus by near infrared spectroscopy (NIRS). *J. Food Qual* **25**: 277-287 (2002)
- ²⁹ Martínez-Vega M, Sharifzadeh S, Wulfsohn D, Skov T, Clemmensen L, Toldam-Andersen T. A sampling approach for predicting the eating quality of apples using visible-near infrared spectroscopy. *J Sci Food Agric* **93**: 3710-3719 (2013)
- ³⁰ Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron K, Lammertyn J, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy : A review. *Postharvest Biol. Technol* **46**:99-118 (2007)
- ³¹ Suykens JAK, Vanderwalle J, Least squares support vector machine classifiers. *Neural process. Lett.* **9**:293-300 (1999)
- ³² Sun T, Lin H, Xu H, Ying Y, 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 Biol. Technol* **51**: 86-90 (2009)
- ³³ Chauchard F, Cogdill R, Roussel S, Roger JM, Bellon-Maurel V, Application of LS-LVM to non linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemom. Intell. Lab. Syst* **71**: 141-150 (2004)
- ³⁴ Williams PC, Implementation of near-infrared technology. In: Williams, P.C.Norris, K.H. (Eds.), Near-infrared Technology in the Agricultural and Food Industries. AACC, Inc., St. Paul, MN,145–169 (2001)
- ³⁵ Cover TM, Thomas JA. *Elem. Inform. Theory*. Wiley, New York (1991)
- ³⁶ Estévez PA, Tesmer A, Perez CA, Zurada JM, Normalized mutual information feature selection, *IEEE Trans. Neural Netw* **20**: 189–201 (2009)
- ³⁷ Guyon I, Elisseeff A, An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **3**: 1157-1182 (2003)
- ³⁸ Koller D, Sahami M, Toward Optimal Feature Selection. *Proc. Int. Conf. on Machine Learning*, 284-292 (1996)
- ³⁹ Herrera LJ, Pomares H, Rojas I, Verleysen M, Guillén A, Effective Input Variable Selection for Function Approximation. *Lecture notes in computer science, Springer, ICANN P*, **4131**:41-50 (2006)
- ⁴⁰ Kraskov A, Stgbauer H, Grassberger P, Estimating mutual information. *Phys.Rev.,E* **69**, 066138 (2004)
- ⁴¹ Golding, J B, L. Spohr, et al, Nondestructive assessment of peach and nectarine firmness. *Acta Hort. (ISHS)* **687**: 261-270 (2005)
- ⁴² Herrero A, Diezma B, Lleó L, Valero C, Ruiz-Altisent M, Evaluación de técnicas acústicas para la determinación de firmeza en melocotón. *REF cong 2009-92. Agroingeniería* (2009)
- ⁴³ Fu X, Ying Y, Zhou Y, Xie L, Xu H, Application of NIR spectroscopy for firmness evaluation of peaches. *Journal of Zhejiang* **9**: 552-557 (2008)
- ⁴⁴ Costa M, Noferini M, Fiori G, Misericocchi O, Bregoli AM, NIRS evaluation of peach and nectarine fruit quality in pre- and post- harvest conditions. Proc. 5th IS on Peach. Eds. R.S. Johnson & C.H. Chrisosto. *Acta Hort.* **592**, ISHS 2002.

- ⁴⁵ Qing Z, Ji B, Zude M, Wavelength selection for predicting physicochemical properties of apple fruit based on near-infrared spectroscopy. *J Food Qual* **30**:511-526 (2007)
- ⁴⁶ Liu F, He Y, Wang L, Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis. *Anal. Chim. Acta.* **615**:10-17 (2008)
- ⁴⁷ Moons E, Dardenne P, Dubois A, Sindic M, Non-destructive visible and NIR spectroscopy measurement for the determination of apple internal quality. *Acta Hort* **517**: 441-448 (1997)
- ⁴⁸ Tu K, De Busscher R, De Baerdemaeker J, Schrevens E, Using laser beam as light source to study tomato and apple quality non-destructively. *Food Processing Automation IV Conferenc.* Chicago, IL, 3-5: 528-536 (1995)

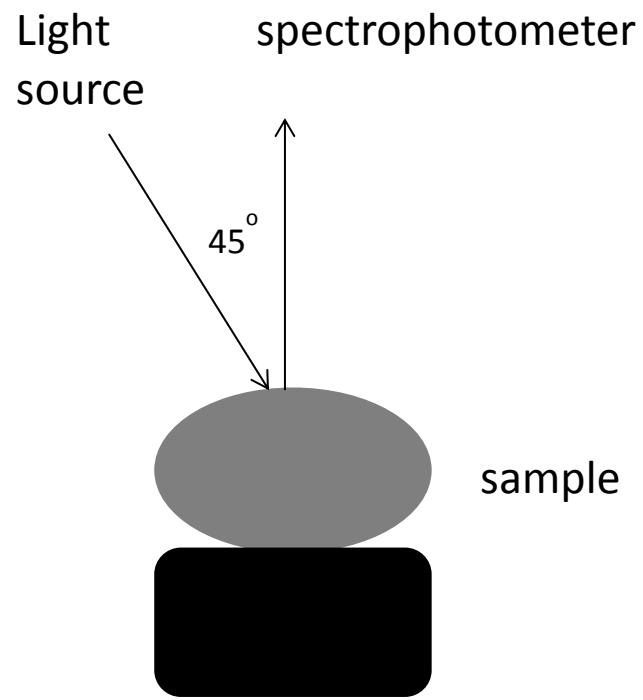


Figure 1. Setup for the acquisition of reflectance spectra

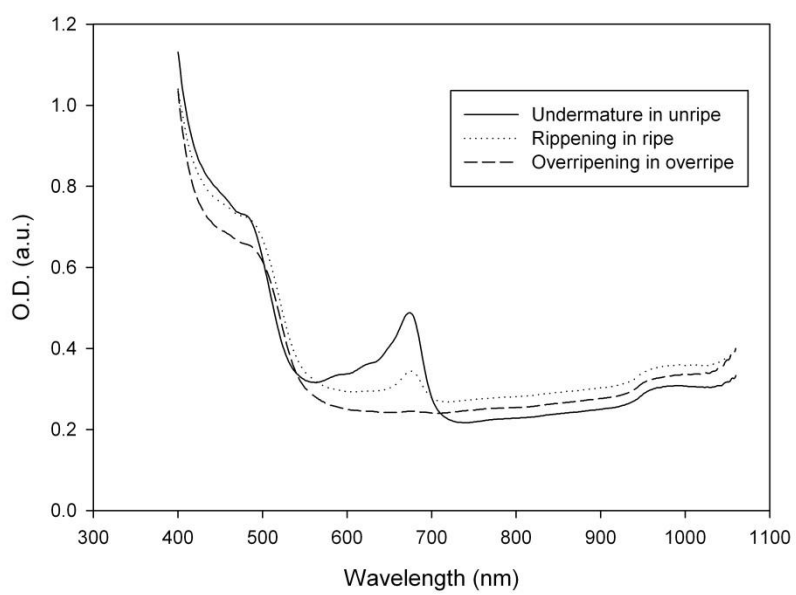


Figure 2. Absorption spectra of peaches in three ripening stages.

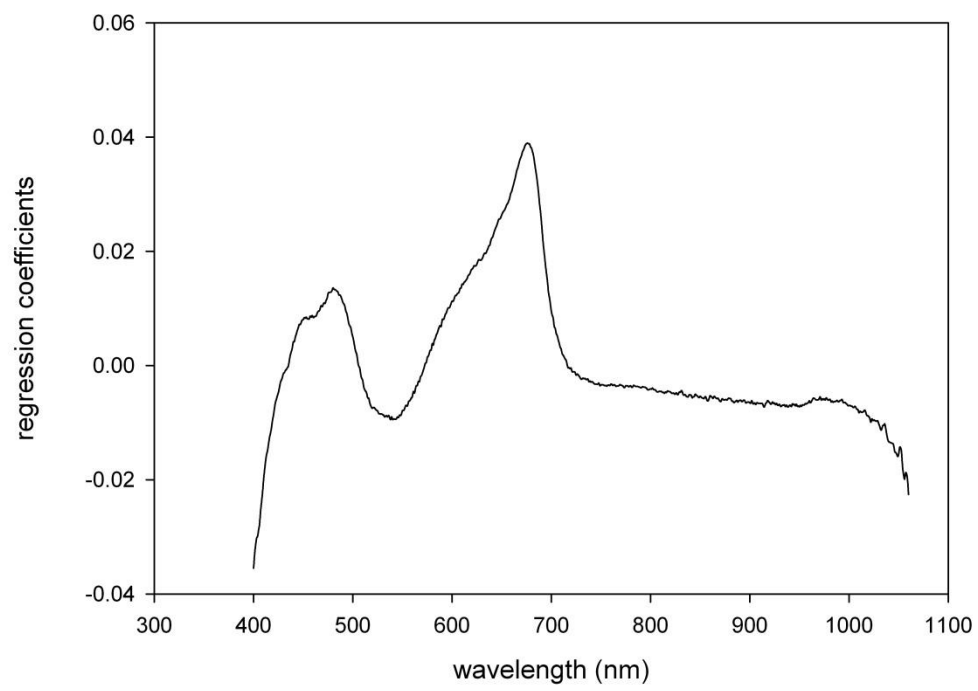


Figure 3. Regression coefficients for firmness for calibration group 1

Variable selection for group 3

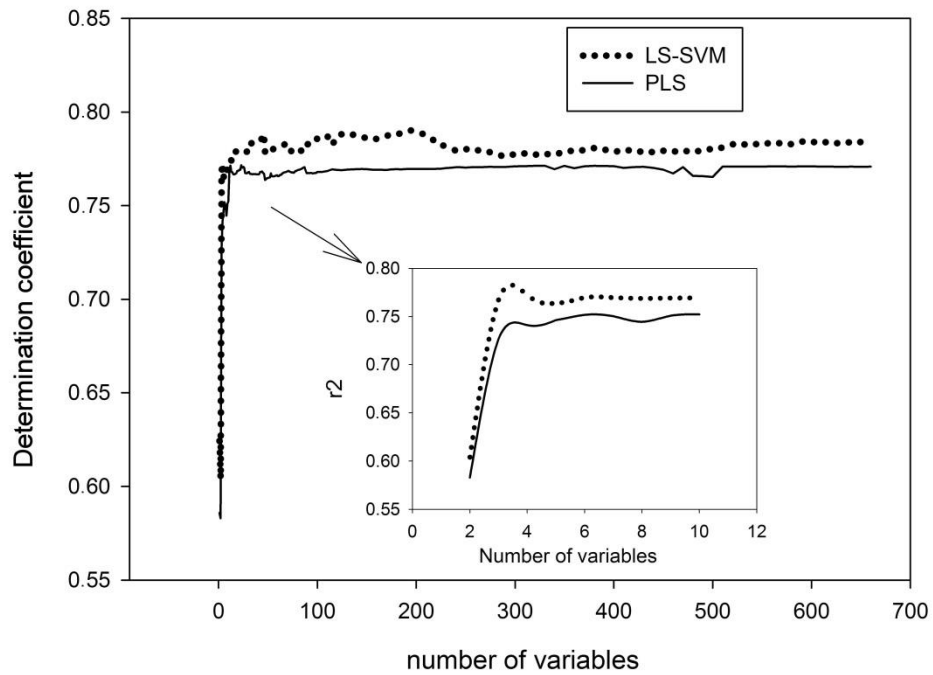


Figure 4. Example of variable selection for both regression models (LS-SVM) red and (PLS) green, using all variables (661) and using only ten variables

Table 1. Samples number (n), mean (X), standard deviation (SD), range and coefficient variation (%CV) for firmness in the three calibration and validation groups

	Group 1		Group 2		Group 3	
	Calibration	Validation	Calibration	Validation	Calibration	Validation
N	194	66	194	66	194	66
X	3.9	3.6	3.7	3.5	3.8	3.7
SD	1.79	1.61	1.7	1.57	1.76	1.7
Range	1-8.7	0.6-8.7	1.2-8.7	1-8.7	0.6-8.7	1.2-8.7
%CV	47.26	44.72	46.21	44.86	46.37	46.41

Table 2. Calibration statistics for firmness using PLS and LS-SVM regression methods for Vis/NIR and Vis/NIR+AWETA. n:number of samples; PLS: number of components; SECV: standard error of cross validation; R_{cv}^2 : coefficient of determination for cross-validation; RPD: residual predictive deviation.

	Vis/NIR			Vis/NIR+AWETA		
	PLS					
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
N	194	194	194	194	194	194
PLS	5	5	5	5	5	5
SECV	0.97	0.99	0.92	0.91	0.92	0.9
R_{cv}^2	0.69	0.71	0.74	0.74	0.74	0.74
RPD	1.84	1.82	1.91	1.97	1.96	1.95
LS-SVM						
N	194	194	194	194	194	194
PLS	5	5	5	5	5	5
SECV	0.89	0.87	0.91	0.83	0.83	0.84
R_{cv}^2	0.64	0.71	0.61	0.86	0.84	0.79
RPD	2.00	2.07	1.93	2.16	2.17	2.09

Table 3. Validation statistics using PLS and LS-SVM regression methods for the prediction for firmness using Vis/NIR, Vis/NIR+AWETA and Vis/NIR+AWETA with three variables. n: number of samples; RMSEP: Root Mean Square Error of Prediction; r^2 : coefficient of external validation; %CV: coefficient of variation.

	Vis/NIR			Vis/NIR+AWETA			Vis/NIR+AWETA (variable selection)		
	PLS								
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
N	66	66	66	66	66	66	66	66	66
RMSEP	0.85	0.84	0.71	0.79	0.81	0.81	0.77	0.9	0.88
r_p^2	0.71	0.71	0.72	0.75	0.73	0.77	0.76	0.66	0.73
Bias	0.04	-0.01	0.07	-0.005	-0.05	0.01	-0.001	-0.13	-0.005
%CV	23.94	24.21	19.29	22.25	23.34	22.01	21.69	26.55	24.04
LS-SVM									
N	66	66	66	66	66	66	66	66	66
RMSEP	0.9	0.92	0.86	0.86	0.84	0.79	0.76	0.84	0.81
r_p^2	0.68	0.74	0.74	0.71	0.71	0.78	0.77	0.71	0.77
Bias	-0.08	-0.07	0.12	-0.1	-0.16	0.05	0.018	-0.13	-0.019
%CV	26.09	27.54	23.56	24.93	25.07	21.18	21.29	24.78	22.13