CARIBBEAN FOOD CROPS SOCIETY

SERVING THE CARIBBEAN SINCE 1963

CARIBBEAN FOOD CROPS SOCIETY

44

Forty Fourth
Annual Meeting 2008

Miami, Florida, USA

Vol. XLIV – Number 2
Plenary Session and Oral Presentations

MEETING HOST:

UF UNIVERSITY of FLORIDA IFAS

2008 Proceedings of the Caribbean Food Crops Society. 44(2):144-154. 2008

Evaluating Hass Avocado Maturity Using Hyperspectral Imaging.

D. Girod[1], J.A. Landry[1], G. Doyon[2], J.A. Osuna-Garcìa[3], S. Salazar-García[3] and R. Goenaga[4] [1]Department of Automated Production Engineering, École de technologie supérieure, Université du Québec, Montréal, Canada. [2]Food Research and Development Centre, Agriculture and Agri-Food Canada, Saint-Hyacinthe, Canada. [3]INIFAP-Santiago Ixcuintla Experimental Station, Nayarit, Mexico. [4]USDA-ARS, Tropical Agriculture Research Station, Mayagüez, Puerto Rico.

ABSTRACT.
The maturity of avocado fruit is usually assessed by measuring its dry matter content (DM), which is a destructive and time consuming process. The aim of this study is to introduce a non-destructive and quick technique that can estimate the DM content of an avocado fruit. 'Hass' avocado fruits at different maturity stages and varying skin color were analyzed by hyperspectral imaging in reflectance and absorbance modes. The DM ranged from 19.8% to 42.5%. The hyperspectral data consist of mean spectra of avocados in the VIS/NIR region, from 400nm to 1000nm, for a total of 163 different spectral bands. Relationship between spectral wavelengths and DM content were carried out using a chemometric partial least squares (PLS) regression technique. Calibration and validation statistics, such as correlation coefficient (R^2) and prediction error (RMSEP) were used as means of comparing the predictive accuracies of the different models. The results of PLS modeling, over several different randomizations of the database, with full cross validation methods using the entire spectral range, resulted in a mean R^2 of 0.86 with a mean RMSEP of 2.45 in reflectance mode, and a mean R^2 of 0.94 with a mean RMSEP of 1.59 for the absorbance mode. This indicates that reasonably accurate models (R^2>0.8) could be obtained for DM content with the entire spectral range. Also this study shows that wavelengths reduction can be applied to the problem. Starting with 163 spectral bands, the DM could be predicted with identical performances using 10% of the initial wavelengths (16 spectral bands). Thus the study demonstrates the feasibility of using VIS/NIR hyperspectral imaging in absorbance mode in order to determine a physicochemical property, namely DM, of 'Hass' avocados in a non-destructive way. Furthermore it gives some clues about which spectral bands could be useful for that purpose.

KEYWORDS: Non-destructive, dry matter, avocado, spectroscopy, imagery

INTRODUCTION
Avocado's (*Persea* Americana Mill.) maturity is indicated by it's oil content and is commercially measured using dry matter (DM) analysis, both of which were shown to be highly correlated (Charles, 1978; Lee et al., 1983). The measurement of both is long and destructive for the fruit. Finding a non-destructive mean of evaluating the maturity of an avocado could prove useful in a production context. Rapid, accurate and easy measurements in the field to monitor produce maturity, allowing for timely harvest would be very valuable, both to the producer (economics) and to the consumer (quality).

The field of spectroscopy is an emerging technology in the agro-food industry and is seeing new applications in the assessment of produce quality and safety. The use of spectrometry may provide fast and reliable methods to this industry and more research is definitely in order.

MATERIAL AND METHODS

Fruits

A total of 21 avocados of the Hass variety were bought from a local market. It is known that there is a DM gradient repartition within the fruit (Woolf et al., 2003). In consequence, dry matter content of fruit was determined by taking thin flesh slices lengthwise, over the 4 cardinal points with respect to the peduncle of the fruit. The slices were peeled and stoned to form 10±1g samples. Dry matter was determined by weight loss following 5 hours drying in a laboratory oven at 105°C.

Hyperspectral imagery

Visible/Near Infrared (VIS/NIR) spectra of intact avocados were collected over the 400nm-1000nm range of the electromagnetic spectrum, with a 2.5nm resolution in reflectance mode, resulting in 163 spectral bands. The hyperspectral imaging system consists of a 12 bits CCD camera (Imperx IPX-2M30), a V10E spectrograph (Specim, FI), a Schneider Kreuznah 23mm achromatic lens, and four (4) 300W Tungsten halogen spots (Ambico V-100). Each reflectance spectrum was then calibrated using dark and white references in order to convert it to reflectance values. The dark reference was obtained by covering the lens with a cap, in order to prevent light from reaching the sensor. The white reference, measured for every 6 fruits, was a Spectralon tile, widely used in spectroscopy (Springsteen, 1999) . Data acquisition and storage was achieved with a PC running in-house software (SpectralCube v2.72, Autovision, USA).

Measurements

In order to match the spectral data to the DM content measured, 4 regions of interest upon the fruit were identified, corresponding to the 4 regions where the DM content was measured. For each region, the corresponding spectrum was obtained by averaging the spectra of this region. Each spectrum is then composed of 163 reflectance values.

Statistical dry matter content analysis

Statistical analysis of the DM database was performed in order to analyze the adequacy of the sampling (number of intra-fruits measures: 4), and repetition rates (number of inter-fruits measures: 21). It is important to select an acceptable variability for a given error level (α), which leads to a sample size and or a number of repetitions suitable for the experiment. Calculating the coefficient of variation (Equation 1), with σ the standard deviation, and $\bar{x}$ the mean, and evaluating what coefficient of variation is desirable, the target (wished) standard deviation can be computed (Equation 2). Knowing the actual (obtained) and the desired (wished) standard deviations for given sampling and repetition rates, the desired (wished) sample size and number of repetitions can therefore be calculated (Equation 3) where t is the student distribution and α the error level (typically 0.05 corresponding to 95% confidence).

$$CV_{obtained} = \frac{\sigma_{obtained}}{\overline{x}_{obtained}} \times 100\ (\%) \qquad \text{(Equation 1)}$$

$$\sigma_{wish} = \frac{\overline{x}_{obtained} \times CV_{wish}}{100} \qquad \text{(Equation 2)}$$

$$N = \left(t_{\frac{\alpha}{2}, N_{obtained}-1} \times \frac{\sigma_{obtained}}{\sigma_{wish}} \right)^2 \qquad \text{(Equation 3)}$$

Modeling the relation between NIR data and dry matter

Analysis of the NIR and DM data involved 50 separate modeling/test exercises. Each exercise required a different random split of the database into 2 subsets: a calibration/modeling set ($n=55$ fruits), and a validation/test set ($n=29$ fruits). The minimum error of cross validation (RMSECV, 5 subsets) was used to choose the number of latent variables in the model, up to a maximum of 40. Once the modeling completed, the calibration model was applied to the test subset, and regression analysis between predictions and known physical measurements was used to judge the predictive abilities of the model. The statistics used were the correlation coefficient R^2, (Equation 4), and the root mean square error (Equation 5) over the calibration set (RMSEP) and validation set RMSEP. In these equations $\hat{y}_i$ is the predicted value by the model and $\overline{\mathbf{y}}$ is the mean value of the actual measurements. Performing 50 different exercises of modeling/testing allowed us to gauge the stability of the models against measurements variations.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{\mathbf{y}})^2} \qquad \text{(Equation 4)}$$

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \qquad \text{(Equation 5)}$$

The principal modeling tool was the Partial Least Squares regression (Matlab R2007b, Mathworks, USA) with the PLS toolbox (v4.1.1, Eigenvector Research, USA). The DM content is modeled as a linear combination of spectral reflectance/absorbance (Equation 7) values weighted by the regression vector coefficients β (Equation 6). Apart from absorbance transformation, no other spectral pretreatments were applied.

$$\%DM = \sum_{i=1}^{163} \beta_i \lambda_i + e \qquad \text{(Equation 6)}$$

$$Absorbance = \log\left(\frac{1}{Reflectance}\right) \qquad \text{(Equation 7)}$$

Band reduction

Variable selection is an important step in the prediction process. In situation where the number of wavelengths (variables) is large, the reduction of this number by using the ones that carry most information may give a safer and easier model to interpret, with fewer factors and better performances. Thus wavelengths not useful to the model are put aside, leaving only the ones that are important. The difficulty then resides in the selection of a relevant variable subset of a reasonable size. The evaluation of all possible models ($2^{163}-1$) being impossible, it is necessary to use a method for the selection of relevant variables.

We selected the PLS approach (Lima, Mello and Poppi, 2005). The pruning approach intends to promote an ordered and selective PLS regression coefficient (variable) elimination based on a specific criterion, trying to reach an improvement in the prediction capability of the model. PLS pruning is an iterative technique of variable elimination. Starting from the total set of variables (163), the aim of the method is the deletion of unimportant PLS spectral bands by minimizing the error variation (saliency calculation) following the deletion of a particular variable. It can be summarized as follows:

(1) Start from a PLS model with the total set of variables (*m=163*)

(2) Eliminate just one variable according to the saliency calculation that estimate the error variation following the exclusion of this band.

(3) Compute a new PLS model with the *m-1* remaining variables.

(4) Execute steps (2) and (3) until there is one variable remaining.

Using this procedure it is possible to collect a set of *m* models, the first one with *m* variables, the second one with *m-1* variables… until the last one with 1 variable. For each model there is an associated error value (RMSECV), the best model is the one that will offer the minimum error with the minimum variables at the same time.

RESULTS

Dry matter measures

DM content values are within 19.8% and 42.5%, with a lack of data around 30% (Figure 1). The coefficient of variation for the sampling number is 2.55%, and 23.36% for the repetitions. A value of 5% is acceptable for the sampling corresponding to a sampling wished equal to 4. But for the repetition a value of 10%, even 5%, would be acceptable; corresponding to a number of repetitions of 24 fruits (10%), and 95 fruits (5%), instead of 21 fruits.

Prediction of dry matter using all spectral bands

Calibration and validation statistics for the DM prediction are listed in Table 1. These are the means over the 50 random subsets of the entire database. The absorbance mode results in better predictive performances with R^2=0.94 and RMSEP=1.65, whereas in reflectance mode we obtained R^2=0.86 and RMSEP=2.41. Typical spectra in absorbance and reflectance modes can be observed on Figure 2.

Figure 1 . Dry matter content sorted by avocado, with 95% confidence intervals

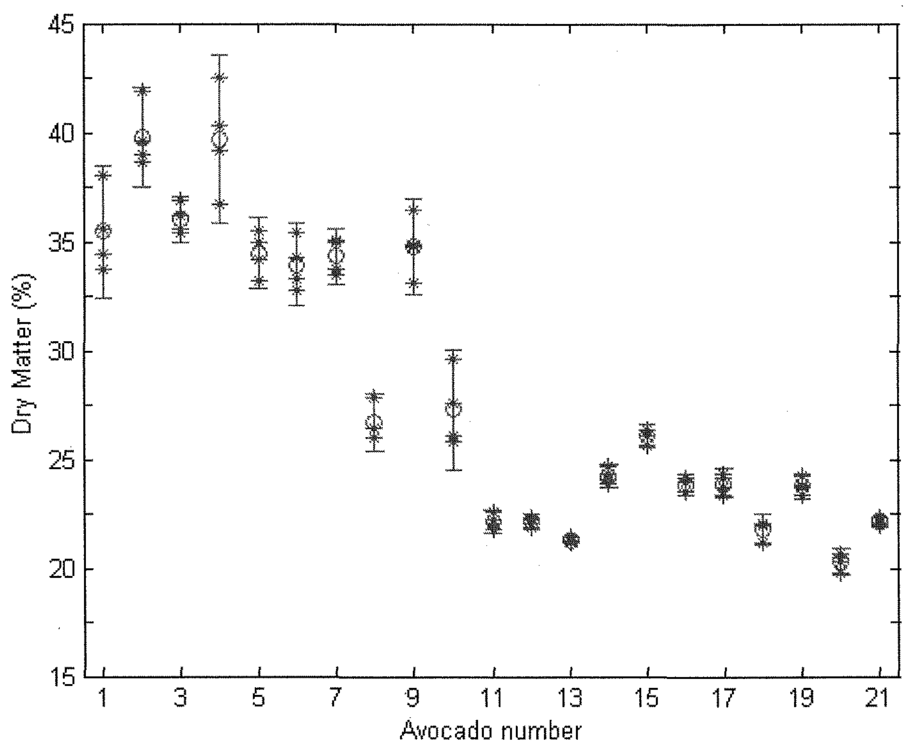
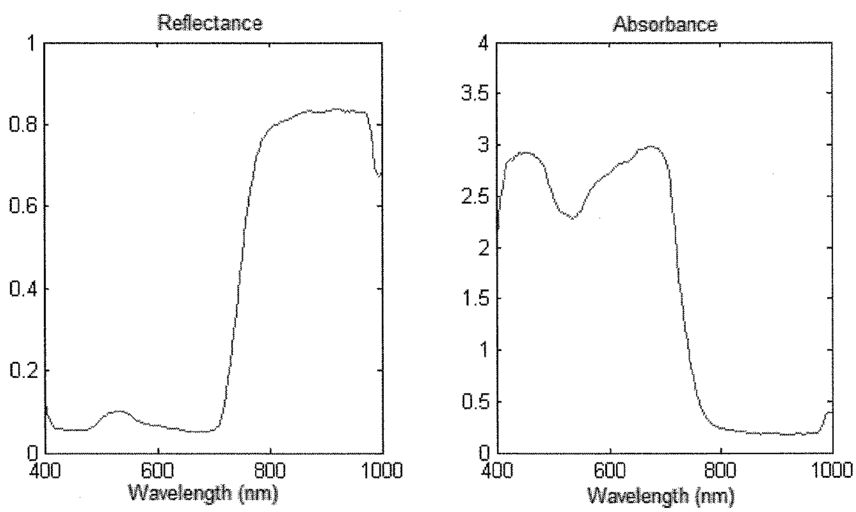


Table 1 . Calibration and validation statistics

	LV	RMSEC	RMSEP	R^2
Reflectance	17	1.10	2.41	0.86
Absorbance	13	0.80	1.65	0.94

The shape of these curves can be explained by the response of the pigments present in the fruit skin. The main pigments responsible for the color of vegetables (skin and flesh) are the chlorophyll (α and β), the anthocyanin and the carotenoid. In the avocado skin, we find mainly both chlorophylls and carotenoïds (lutein and β-carotene) which absorption spectra can be observed on Figure 3 (Ashton et al., 2006). To analyze the differences obtained in the performances between reflectance mode and absorbance mode, the DM data was divided in 4 groups based upon their DM content. The fist group has DM contents inferior to 25%, the second group has DM contents comprised between 25% and 31%, the third one has DM contents between 31% and 37.5%, and the last group has DM content superior to 37.5% (Figure 4). The appearance of mean spectra is conforming to the absorption spectra of the pigments (Figure 5). Poor reflectance around the two absorption peaks at 450nm and 650nm, and after near 700nm at the so-called 'red edge', leading to a plateau of high reflectance in the near-infrared, where pigments no longer absorb radiation (Blackburn, 2007). Comparison of the mean spectra given by those 4 groups shows that the differences between the spectra in the 400-700nm is more important in absorbance mode than in reflectance mode.

Figure 2 . Typical absorption and reflection spectra from avocado



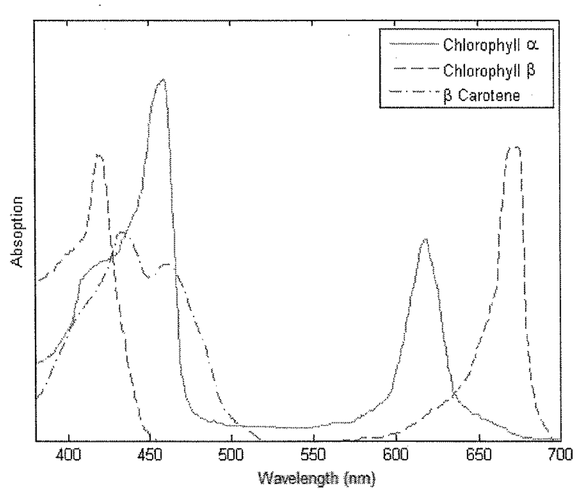


Figure 3 . Absorption

spectra of pigments in avocado skin, adapted from (Ashton et al., 2006)

Band reduction

The PLS pruning method was applied to the 50 random subsets of the database leading to 50 band selection subsets. Therefore, one can compute the number of times each band had been chosen (Figure 6). The 16 bands that obtained the largest number of votes are listed in

Table 2. 14 of these bands are in the visible range (400nm-700nm), which seems to confirm the pigment analysis. With those 16 bands, the most interesting combination can be determined by computing the performances given by all possible combination of those 16 bands (Table 3).

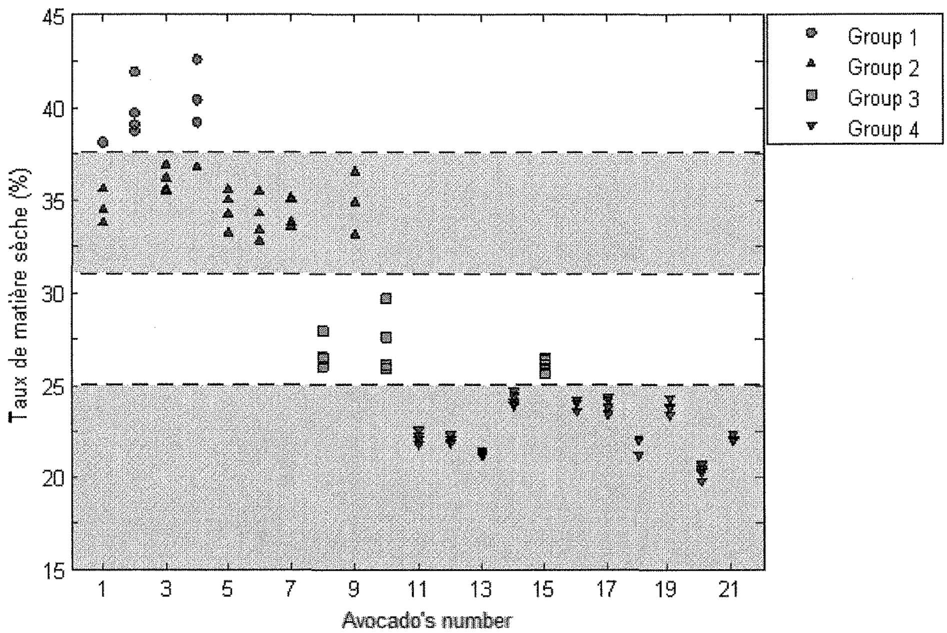


Figure 4 . DM content groups

Band reduction is possible and can even offer better performance than while using all bands. Modeling with 16 bands (10% of the initials set) resulted in R^2=0.96 and RMSEP=1.35. Using as few as 8 bands (5% of the initials set) the performances are equal to that of using the full number of bands (R^2=0.94, RMSEP=1.60). Finally, with only 5 bands (3% of the initials set) performances are still very acceptable (R^2=0.90, RMSEP=2.04). The retained bands can be observed on Figure 7.

DISCUSSION

Using hyperspectral imagery to evaluate the maturity of fresh avocado is possible. There is a strong correlation between the dry matter content of Hass avocado, and its skin color (as perceive by spectroscopy). As there is still uncertainties concerning the origin, transport and storage of the fruits between their harvest and the measurements in the laboratory, as well as a lack of DM samples around 30%, conclusion of this study cannot be definitive, but can lead to effective recommendations for future work.

Figure 5 . Mean spectra of each of the 4 avocado groups

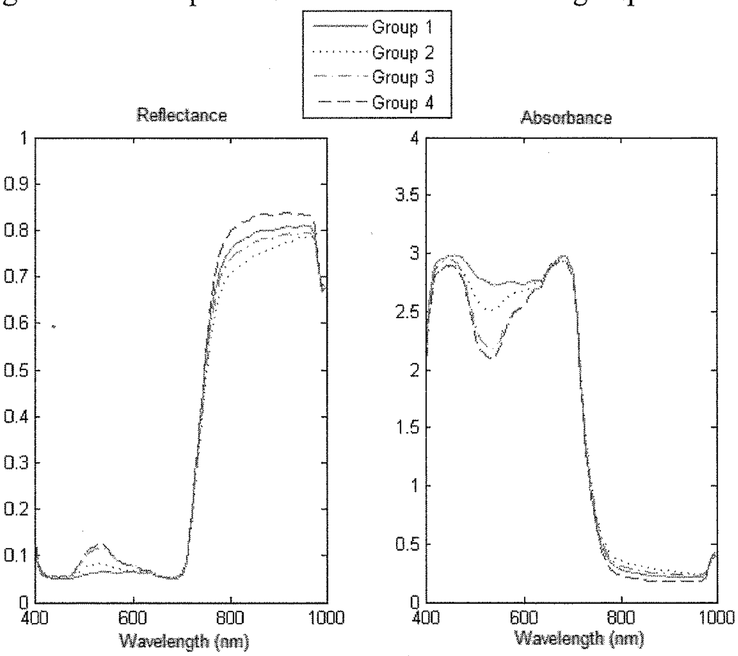


Figure 6 . Pruning votes

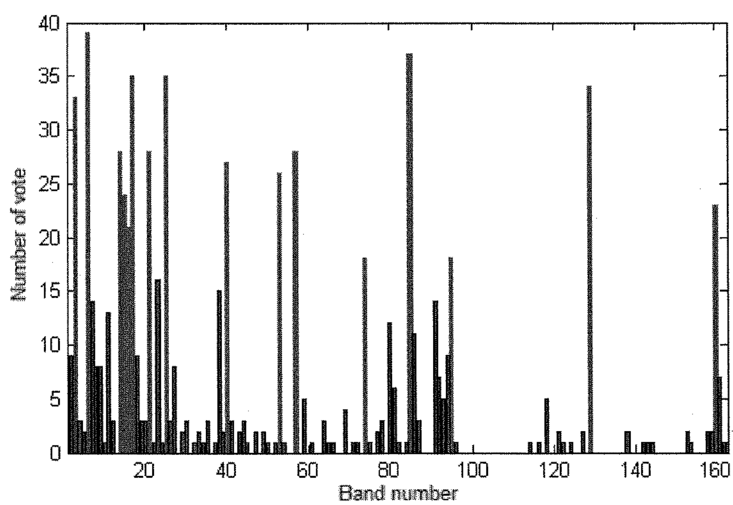


Table 2 . Selected bands and their wavelengths

Selected bands	3	6	14	15	16	17	21	25
Wavelength (nm)	425.58	458.97	486.35	489.27	492.19	495.110	506.88	510.69
Selected bands	40	53	57	74	85	95	129	160
Wavelength (nm)	563.08	601.40	613.16	663.09	695.41	724.83	825.74	959.81

Table 3 . Mean performance of band combinations

Number of bands	RMSEP	R^2	Wavelength of selected bands
5	2.04	0.90	40 57 85 95 129
8	1.60	0.94	16 21 25 40 57 85 95 129
16	1.35	0.96	3 6 14 15 16 17 21 25 40 53 57 85 95 129 160

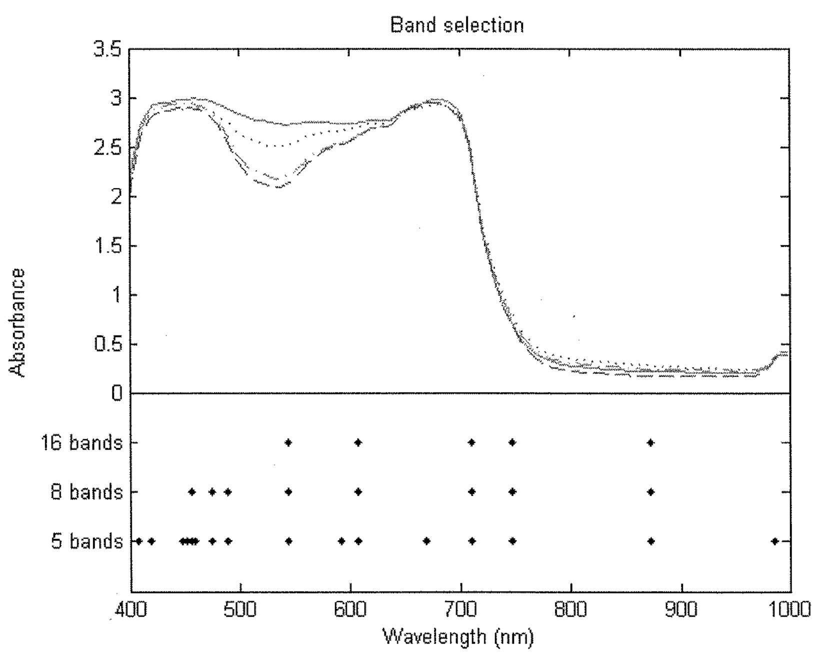


Figure 7 . Band selection

It is necessary to have maximum variability along with an adequate DM content repartition over the dataset. To accomplish that, one need to have a reliable avocado source to satisfy the same condition for storage and transport between harvest and analysis. On site study would be preferable. A study with 24 fruits can lead to a variability coefficient of 10%, if the goal is 5%, 95 fruits are needed. Measuring DM content of each avocado over the 4 cardinal points with respect to the peduncle of the fruit, combined to the acquisition of 4 spectra of the same areas will ensure that the DM content variability is respected.

Following those recommendations, a full study can be carried on, leading to more conclusive results.

REFERENCES

Ashton, O. B. O., M. Wong, T. K. McGhie, R. Vather, Y. Wang, C. Requejo-Jackman, P. Ramankutty et A. B. Woolf. 2006. « Pigments in Avocado Tissue and Oil ». *J. Agric. Food Chem.,* vol. 54, n° 26, p. 10151-10158.

Blackburn, George Alan. 2007. « Hyperspectral remote sensing of plant pigments ». *Journal of Experimental Botany,* vol. 58, p. 855-867.

Charles, E. Lewis. 1978. « The maturity of avocados - a general review ». *Journal of the Science of Food and Agriculture,* vol. 29, n° 10, p. 857-866.

Lee, S. K. , R. E. Young, P. M. Schiffman et C. W. Coggins. 1983. « Maturity Studies of Avocado Fruit Based on Picking Dates and Dry Weight ». *Journal of the American Society for Horticultural Science,* vol. 108, n° 3, p. 390-393.

Lima, Silvio L. T., Cesar Mello et Ronei J. Poppi. 2005. « PLS pruning: a new approach to variable selection for multivariate calibration based on Hessian matrix of errors ». *Chemometrics and Intelligent Laboratory Systems,* vol. 76, n° 1, p. 73-78.

Springsteen, Art. 1999. « Standards for the measurement of diffuse reflectance - an overview of available materials and measurement laboratories ». *Analytica Chimica Acta,* vol. 380, n° 2-3, p. 379-390.

Woolf, Allan, Chris Clark, Emma Terander, Vong Phetsomphou, Reuben Hofshi, Mary Lu Arpaia, Donella Boreham, Marie Wong et Anne Whit. 2003. « Measuring avocado maturity; ongoing developments ». *The Orchadist*, p. 40-45.