THE JOURNAL OF
AGRICULTURAL SCIENCE

CAMBRIDGE
UNIVERSITY PRESS

# AUTOMATIC DISCRIMINATION OF GRAPEVINE (Vitis vinifera L.) CLONES USING LEAVES HYPERSPECTRAL IMAGING AND PARTIAL LEAST SQUARES



| | |
|---|---|
| Journal: | *Journal of Agricultural Science* |
| Manuscript ID: | AGS-2013-00252.R1 |
| Manuscript Type: | Climate Change & Agriculture Research Paper |
| Keywords: | Hyperspectral image , classifier, grapevine cultivar |
| | |

# AUTOMATIC DISCRIMINATION OF GRAPEVINE (*Vitis vinifera* L.) CLONES USING LEAVES HYPERSPECTRAL IMAGING AND PARTIAL LEAST SQUARES

**Abstract**

A worldwide innovative method to discriminate grapevine clones is presented. It is an alternative to ampelography, isozyme and DNA analysis. The spectra and their first and second derivatives of 201 bands in the visible and near-infrared wavelength range between 634 nm and 759 nm were used as input to a classifier created using Partial Least Squares. The spectra were acquired in laboratory for the adaxial side of the apical part of the main lobe of fully hydrated grapevine leaves. The created classifier allowed separating 100 leaves of the Cabernet Sauvignon (*Vitis vinifera* L.) variety into four clones, namely clone CS 15, CS 169, CS 685 and CS R5, comprising 25 leaves each. The percentage of correctly classified leaves of these clones was 98.2, 99.2, 100 and 97.8%, respectively, when the classifier input was the second derivative of the normalized spectra. These percentages were determined by Monte-Carlo cross-validation. With the new method proposed each leaf of a given variety can be classified in a few seconds according to its clone in an environmentally friendly way.



## 1. Introduction

Grapevine varieties are long-living, woody perennial plants, which are asexually propagated. Vegetative propagation for many generations allows the appearance and accumulation of somatic mutations which are the base for the selection of new clones that may exhibit differences in various characteristics such as vigour, berry and cluster weight, yield production, resistance to plagues and diseases and enological potential, among others.

Tools for distinguishing grapevine clones could be very useful to ensure quality management and tracking of grapevine propagation material (Blaich *et al*. 2007). Clone discrimination was initially attempted using descriptive ampelography (Galet 1979) which relies on plant morphological features, as well as isozymes, which target

biochemical traits (Schlapfer & Fischer 1998), without much success. Nowadays, the state-of-the-art methods for clone fingerprinting are based on DNA analysis procedures such as Simple Sequence Repeats (SSR) (Regner *et al.* 2000), Amplified Fragment Length Polymorphism (AFLP) (Cervera *et al*. 2000) or Random Amplification of Polymorphic DNA (RAPD) (Regner *et al.* 2000). The discrimination power of these techniques has been very variable in the literature with positive (Cervera *et al,* 2000; Imazio *et al,* 2002; Fanizza *et al*. 2005; Forneck *et al*. 2009; Blaich *et al.* 2007; Cretazzo *et al.* 2010; Anhalt *et al.* 2011) and negative results (Moncada & Hinrichsen 2007; Konradi *et al.* 2007) depending on the variety and methodology used. In addition, only in specific occasions, microsatellites, which are widely used for variety identification, have also been capable of discriminating grapevine clones (Silvestroni *et al.,* 1997; Kozjak *et al.* 2003; Regner *et al,* 2006). These DNA based methods present problems of repeatability among laboratories and are complex, slow and time-consuming, preventing their use in large scale. These drawbacks create the necessity to develop new methods that mitigate these problems in order to make clone discrimination easy and accessible to grape growers. Among the possible new methods to address grapevine clone discrimination, the combination of spectroscopy and machine learning techniques, which involves computational processes of discovering patterns in large data sets, is a promising and innovative approach.

Spectroscopy is the analysis of how electromagnetic radiation interacts with matter at different wavelengths. Hyperspectral imaging is a spectroscopic technique that allows gathering information for several hundred wavelength bands on various spatial positions. The radiation that reaches the leaf surface may afterwards be reflected, absorbed or transmitted. The nature and fractions of reflection, absorption and transmission depend on the wavelength of radiation, angle of incidence, surface roughness as well as optical properties and biochemical contents of the leaves (Vogelmann 1993 ). In the present work only reflectance was used.

The spectral reflectance properties of plant leaves are determined by the absorption and scattering events occurring within the leaf (Grant 1987). In the visible range (VIS: 400-700 nm), leaf pigments are the main factor responsible for leaf reflectance, whereas in the near infrared (NIR: 700-1350 nm) and short-wave infrared (SWIR: 1400-3000 nm) ranges, the reflectance signature of leaves is determined by structural processes (leaf morphology) and leaf water content (Peñuelas *et al.* 1993), respectively. The transition from the low reflectance at VIS to the high reflectance at NIR is called the red-edge

(Collins 1978). Intrinsic differences in the leaf content of pigments among clones of a given species have been reported in previous works (Taylor *et al*. 1992; Conforto *et al*. 2011; Lin *et al*. 2011). Similarly, a histological study of leaves of several clones of cv. Albariño *(Vitis vinifera* L.) evidenced significant differences in the characteristics and thickness of their spongy mesophyll (Alonso-Villaverde *et al.* 2011). In this way, the distribution of air spaces and the arrangement and size of cells in the mesophyll have a strong impact on NIR reflectance (Gausman *et al*. 1970). Leaves having a compact mesophyll layer with fewer air spaces would allow more transmission and less radiation scattering, and would lead to increased NIR reflectance from the adaxial leaf surface (Terashima & Saeki, 1983; Slaton *et al*. 2001).

Partial least squares (PLS) is a machine learning regression and classification method. Machine learning is the discipline that studies the development of algorithms that provide machines the ability to automatically learn from new examples. Machine learning and spectroscopy have been already used separately for clone determination. Likewise, machine learning algorithms were used to distinguish clones from Eastern Cottonwood (Pande 2011), grapes (Ferrandino & Guidoni 2010) or cocoa (Engels 1983) using chemical or morphological features measured through complex and slow methods. Spectroscopic measurements in their turn, were used to distinguish clones of rubber plant (Osman *et al.* 2010) and jojoba (Gayol *et al.* 2009). In these works, the discrimination power of machine learning was not necessary because clone separation was based on a single feature, such as the maximum reflectance values of the plant seeds, in the case of rubber (Osman *et al*. 2010), or the concentration of a specific chemical compound in the case of jojoba (Gayol *et al.* 2009). The maximum reflectance values are not discriminative for grapevine clones and the Gayol *et al.* method requires a conventional calibration to determine the chemical compounds concentration, calibration which is avoided with the new method proposed.

The present work proposes an innovative alternative to ampelography, isozyme and DNA analysis for grapevine clone discrimination, consisting of the combination of spectroscopic (hyperspectral imaging) and machine learning algorithms (PLS) to classify grapevine clones using the leaves. PLS was used to create the clone classifier, i.e. the mathematical function that transforms the hyperspectral data into a given "clone type identification". The proposed method aims at discriminating grapevine clones by combining the processing and analysis power of machine learning algorithms with the simplicity of gathering sample information using portable spectroscopes.

## 2. Methods and Materials

### *2.1. Leaf samples*

Leaves of grapevines of *Vitis vinifera* L. cv. Cabernet Sauvignon were collected at the 5th of October 2011. A number of 100 leaves from four clones named CS 15, CS 169, CS 685 and CS R5 were collected in the field (Vitis Navarra nursery, Tafalla, Spain) and Bodegas Pago de Larrainzar (Estella, Spain) (Table 1). The leaves, 25 per clone, were collected from 48 different vines, 12 vines per clone, at various maturity stages or shoot node positions, to have a representative set of leaf samples per clone. After being collected, the leaves were immediately submerged in deionized water at 10ºC. This was intended to avoid desiccation as well as to bring all the leaves to a similar water status corresponding to fully hydrated leaves with approximately 100 % relative water content. The leaves were then transported in portable refrigerators to the University of Rioja and were kept in cold water for 18 hrs. After this time, each leaf was blotted dried and a disc of 2 cm diameter was cut in the apical part of the main lobe. Hyperspectral images of the leaves discs were immediately taken on the adaxial side of the fully hydrated leaves discs.

### *2.2. Experimental setup and image acquisition*

The experimental setup, see Figure 1, was composed of a hyperspectral camera, lighting and a white reference made of Spectralon® (Specim, Oulu, Finland). Spectralon® has the highest diffuse reflectance of any known material or coating over the ultraviolet, visible, and near-infrared regions of the spectrum, allowing to measure the total intensity of light that was incident on the leaf surface. The hyperspectral camera consisted of a JAI Pulnix black and white camera (JAI, Yokohama, Japan), that registered the image, and of a Specim Imspector V10E (Specim, Oulu, Finland) spectrograph that decomposed the light into its composing wavelengths. The camera comprised also a lens. The black and white camera image size was 1040 by 1392 pixels. The 1040 pixels corresponded to the same number of wavelength bands from 380 nm to 1028 nm, with each band having approximately 0.6 nm width. The 1392 pixels were relative to the spatial dimension and each corresponded to a width of 79 μm.

In the present work, only the visible and part of the near-infrared regions, between 634 and 759 nm, and corresponding to 201 bands, were used. The wavelength region used was strongly narrowed to facilitate the potential development of cheap dedicated

equipments after this work. Nevertheless, the narrowing could not be too extreme or it would prevent the development of good classifiers due to discarding important information. The region was chosen to include the low red chlorophyll reflectance and the red-edge which constitute important features of the spectral response of vegetation.

For each leaf sample 30 hyperspectral images were taken. Image acquisition was done using Coyote software (Version 2.2.0, JAI, Japan) at a frequency of eight images per second. The hyperspectral camera was placed at 420 mm from the sample. Its spectral calibration was done using the manufacturer procedure (EHE 2006) and the spatial calibration using a black and white target. The lighting was done using four 20W, 12V halogen lamps and two 40W, 230V reflector lamps from Philips (Spotline, Philips, Eindhoven, Nederlands). The lamps were chosen to provide the best possible illumination over the wavelength operation range of the hyperspectral camera. They were powered by direct current power supplies to avoid image flickering. The dimensions of the lamps holder were 300 x 300 x 175 $mm^3$ (length x width x height). The measurements were done in a dark room at 20 degrees Celsius.

### *2.3. Reflectance determination*

The experimental setup allowed to determining the samples' reflectance (R). The reflectance spectrum provides the percentage of light that is reflected by a leaf relative to the total intensity of light that is incident on the leaf for each wavelength (λ) measured. In hyperspectral images the reflectances were determined for each spatial position $x$. Reflectance for a certain position $x$ and wavelength λ was determined according to equation 1:

$$R(x,\lambda) = \frac{LI(x,\lambda) - D(x,\lambda)}{SI(x,\lambda) - D(x,\lambda)} \quad \text{(eq. 1)}$$

where $LI(x,\lambda)$ is the intensity of light coming from the leaf, $SI(x,\lambda)$ corresponds to the total light intensity that reaches the leaf, and $D(x,\lambda)$ is the camera's dark current. $SI(x,\lambda)$, the reference for reflectance estimation, was measured using the white reference made of Spectralon®. The dark current was measured with the camera lens covered and corresponds to inherent electronic noise. Its value must be discounted from both the leaf and Spectralon® signals because it is independent from these signals and

would distort the real reflectance values. The parameters $LI(x,\lambda)$, $SI(x,\lambda)$ and $D(x,\lambda)$ were calculated from averaging the 30 hyperspectral images per leaf sample to minimize measurement noise. The reflectance values were averaged over all the spatial positions measured on the adaxial sides of the leaf discs. Finally, the values were normalized to the range between zero and one, originating *NormR*. This way, clone discrimination was based on spectra shape and not on absolute reflectance values. As an input to the classifiers, besides *NormR*, it was also used the first derivative (*NormR*$^{(1)}$) or the second derivative (*NormR*$^{(2)}$) of the normalized reflectance. The two derivatives were calculated using the following equations:

$$NormR^{(1)} = \frac{-NormR_{i+2}+8\,NormR_{i+1}-8\,NormR_{i-1}+NormR_{i-2}}{12} \quad \text{(eq. 2)}$$

$$NormR^{(2)} = \frac{-NormR_{i+2}+16\,NormR_{i+1}-30\,NormR_{i}+16\,NormR_{i-1}-NormR_{i-2}}{12} \quad \text{(eq. 3)}$$

Where *i* is the wavelength at which derivatives are being calculated and *i+1* refers to the hyperspectral camera wavelength band after *i*.

### *2.4. Partial Least Squares*

Partial least squares (PLS) was used to develop the classifiers that provided a clone class for hyperspectral reflectance spectra of leaves. This method constructs a model that transforms independent variables, in the present case, the leaves reflectance values, or the reflectance derivatives, into a dependent variable, the clone class. PLS works by transforming the independent variables into a new set of variables that explain to the largest possible extent the covariance between independent and dependent variables (Geladi & Kowalski 1986). A small percentage of new independent variables, relatively to the number of initial variables, is usually able to account for most of the variability in the data. The final step consists of a regression providing the coefficients that allow transforming the independent variables into the dependent ones. The number of new variables, also called components, varied between one and 79 for each classifier to determine the optimal value which corresponded to the lowest root mean squared error (RMSE) for a validation set.

Quaternary classifiers were created to discriminate the Cabernet Sauvignon *(Vitis vinifera* L.) leaves, meaning that a single classifier was capable of discriminating the four clones under study. The classifiers inputs were the raw, first derivative and second

derivative of the normalized reflectance. The classifiers had four output variables, one for each clone to be identified. Ideally, for a leaf of a certain clone, the classifier output variable assigned to the clone had value one and the remaining variables had value zero. A leaf was attributed to the clone whose corresponding output variable presented the largest value.

All calculations were done using Matlab (Release 2010a, MathWorks, Natick, USA).

### *2.5. Validation*

The validation process aimed at verifying that the PLS model was capable of providing good results for data that were not employed in its creation, i.e. that the model is able to generalize. In the present work the validation method employed was Monte-Carlo cross-validation where multiple training and validation sets were randomly selected from the available data (Arlot 2010). A PLS model was developed using each pair of training and validation sets; each of these PLS models is called a repetition. In each repetition, a certain leaf spectrum cannot be assigned to both training and validation sets. The final validation result corresponds to the average of the validation results for all repetitions. This method was used due to the natural variability of the leaf features which originated different training and validation results as the training and validation sets composition varied.

The validation of the different types of classifiers' was done using 100 Monte-Carlo cross-validation repetitions per classifier type. Each repetition used 80 leaves for training and 20 for validation, with each one of the four clones contributing with 20 leaves for training and five leaves for validation.



## 3. Results

Clone discrimination was based on the differences in the shape, variation and curvature of the normalized reflectance curves. The variation and curvature were calculated using the first and second derivative of the normalized reflectance, respectively. The minimum and maximum values of the normalized reflectance spectra and of the first and second derivatives of these spectra are depicted in

Figure 2a, b and c, respectively. For practically all the cases shown, the maximum reflectance values for the different clones' spectra were almost coincident. The same

happened with the minimum values. Consequently, there were no strong differences between clones' spectral signatures, but overlapping among their spectra. Only in the case of the normalized reflectance of CS 15 the maximum values appeared considerably different from those of all other three clones, but the overlapping still existed. The red-edge was determined for each clone as the maximum values of the first derivative of the non-normalized reflectance spectra in the wavelength region below the 800 nm. It was found at 758 nm, 753 nm, 745 nm and 754 nm for clone CS 15, CS 169, CS 685 and CS R5, respectively.

Figure 3 depicts the distribution of percentages of correctly classified leaves in the validation set of 100 Monte-Carlo repetitions. The results are shown for three different classifier inputs, namely, raw, first derivative and second derivative of the normalized reflectance which are designated by the prefixes "r", "d1" and "d2", respectively. A total of twelve box plots are shown in Figure 3, four per classifier, with each box plot corresponding to one of the clones to discriminate. Table 2 shows the average results of the Monte-Carlo repetitions from Figure 3. These average results were the final validation results of the classifiers. The results in Figure 3 and Table 2 showed an improvement in the classification results when the classifier input changed from normalized reflectance to the first derivative and then to the second derivative. Only in the case of CS R5 the classification percentage was better for the first derivative of the normalized reflectance ("d1-CS R5") than for the second derivative ("d2-CS R5"). Overall, the best classifier input was the second derivative of the normalized reflectance because it allowed obtaining the best mean classification percentage for the four clones. The mean classification percentage for each clone improved as the number of repetitions with correct classification percentages smaller than 100% was reduced. Per example, from the twelve box plots, only "r-CS 169" and "r-CS 685" presented 25$^{th}$ percentiles of 80%; for "d1-CS 169" and "d1-CS 685", which showed better mean classification percentages than "r-CS 169" and "r-CS 685", the 25$^{th}$ percentile reached 100%. Another example involved the classifiers whose input was the normalized reflectance. Their 5$^{th}$ percentile whiskers were at a classification percentage of 60% while this percentile was equal to or larger than 80% for the first and second derivative of the normalized reflectance, which presented better mean results. Using the same rationale it is possible to conclude that the decrease in correct classification percentage from "d1-CS R5" to "d2-CS R5" was caused by a decrease of 5$^{th}$ percentile whisker from 100% down to 80%. The classification percentage median, indicated by the grey

arrows, was 100% for all box plots, meaning that 50% of the repetitions yielded correct classification results of 100%.The best classifier input allowed obtaining correct classification percentages of 98.2%, 99.2%, 100%, 97.8% for the leaves from clones CS15, CS 169, CS 685 and CS R5, respectively.

Figure 4 shows the distribution of the root mean squared error (RMSE) values obtained in the validation sets for the 100 repetitions of the Monte-Carlo cross validation. Results are shown only for the classifiers with the best input, the second derivative of the normalized reflectance. Figure 4 shows four box plots, each one corresponding to an output variable of the classifier and to the clone associated with the variable. The correspondence between output variable and clone is pertinent because, per example, when the first output variable presented a value close to one there was a high probability that the leaf belonged to clone CS 15. From these results it could be observed that it was harder to determinate if a leaf belonged to clones CS 169 or CS 685 than to clone CS 15. In addition, CS R5 was the clone to which leaf allocation was hardest, as the RMSE mean value for CS R5 was the largest and close to 0.21. The 75$^{th}$ percentile of the RMSE was smaller than 0.24 for all box plots and the largest RMSE was smaller than 0.3; these values are relatively small when considering that the optimal values, of the classifier output variable, for belonging or not to a clone are one and zero, respectively. Consequently, the classifier output values were relatively far from the midway value of 0.5 which could lead to unreliable classification.

Figure 5 shows the variance of the second derivative of the normalized reflectance and of the clone classes, in the whole sample set, that is explained by the PLS components. Considering the second derivative, with twenty components, only 59% of the variance was explained. The amount of variance explained increased slowly with the number of components. With respect to the clone class, the same twenty components allowed to explain 98.5% of the variance. Only three components explained 80.9% of the variance and it was necessary nine components to explain 94.1%. For more than nine components the curve increased at a lower rate than for less than nine components.

The distribution of the best number of components in the 100 Monte-Carlo repetitions carried out for the classifiers with the best input, the second derivative of the normalized reflectance, is presented in the histogram of Figure 6. The mean value of components was nine and this value had the largest frequency, however, seven components presented only a slightly smaller frequency. There were no classifiers with more than fourteen components. With nine components, only 36.9% of the variance of the second

derivative of the normalized reflectance was explained. However, this unusually small percentage was enough to create classifiers that presented large correct classification percentages. The small number of components relative to the number of training samples, which was 80, helped the classifier to have good generalization ability.

Figure 7 depicts the loadings of the PLS components of the whole data set when the input was the second derivative of the normalized reflectance. These loadings show the importance of each wavelength to the classification model. Figure 7 shows the loadings of nine components which was the average value of components employed by the classifiers. The variation in loading values from one wavelength to the next was most probably caused by the variation in the values of the second derivative of the normalized reflectance. In the first and second components (Figure 7a, b) there were peaks over the whole wavelength range whereas in the third component (Figure 7c) the loadings for wavelengths smaller than 720 nm became smaller than the loadings for wavelengths larger than 720 nm. This tendency evolved more evident for the loadings of components four to nine (Figure 7d-7i). Nevertheless, mainly in the first three components, the whole range between 634 nm and 759 nm was employed in the components formation and was relevant to the classification model.



**4. Discussion**

Leaf spectral reflectance data, which contains information regarding differences in leaf traits, has been employed in species discrimination (Castro-Esau *et al.* 2004; Asner *et al*. 2008; Sánchez-Azofeifa *et al*. 2009; Burkholder *et al*. 2011; Durgante *et al.* 2013) and may also potentially contribute to clone differentiation within a given genotype. In grapevines, Lacar *et al.* (2001) and Ferreiro-Armán *et al.* (2006) showed the feasibility of spectral imaging on leaves to successfully discriminate among grapevine varieties. Varietal discrimination of grapevines by spectroscopy is an important achievement, but the possibility of clone discrimination within a given grapevine variety from spectral imaging of leaves would be one step forward. In the present study the discrimination of a Cabernet Sauvignon clone based on leaf spectroscopy in the visible and near-infrared range, among a group of four different clones was achieved with a percentage of correct classification which exceeded 97.8 % for the four tested clones, CS 15, CS 169, CS 685 and CS R5.

The wavelength range employed for the classifiers creation was strongly reduced (between 634 and 759 nm) to facilitate the potential subsequent development of cheap dedicated equipments for clone discrimination at a commercial level. This is possible because the narrow range allows reducing equipment complexity and the chosen wavelength values may facilitate the development of good lighting at low cost as well as the use of efficient and cheap light sensors. The wavelength range used covered part of the red reflectance region, characterized by large absorption by chlorophylls (from 634 to 690 nm) and the red-edge, typically occurring at wavelengths between 690-740 nm depending on leaf surface variability (Dockray 1981). The use of the chlorophylls absorption region was pertinent because intrinsic differences in the leaf content of pigments among clones of a given species have been reported in previous works in tea plant (Taylor *et al.* 1992), rubber tree (Conforto *et al.* 2011) and camphor tree (Lin *et al.* 2011). The red-edge was an interesting region because it was where the normalized reflectance values presented the largest variance, when the whole wavelength range between 380 and 1028 nm was considered. The first and second derivative of the normalized reflectance also presented considerable variances at the red edge. In fact, the loadings for classifiers with the second derivative of the normalized reflectance as an input were larger at the red-edge region than in the remaining wavelengths in the range between 634 and 759 nm. The red-edge position varied for the four Cabernet Sauvignon clones studied, being found between 745 nm and 758 nm, probably because increased chlorophyll concentration shifted the red-edge to longer wavelengths (Dockray 1981). In view of all this, the range used in classifiers creation put focus on the influence of chlorophyll over the reflectance spectra. However, since the transition at the red-edge from chlorophyll absorption to within leaf scattering is not abrupt, probably, the developed classifiers also considered some information regarding the scattering.

Some of the identified sources of spectral variability include leaf age, adaxial and abaxial leaf surfaces, sun exposure, and water content (Ribeiro da Luz 2006). From these, the first three factors would be more important in the visible range than the leaf water content. However, with the aim of avoiding any interference from differential leaf hydration status, all sampled and imaged leaves were taken to full hydration, and since all images were taken from the adaxial sides, no spectral variability due to the leaf surface side was expected. Regarding leaf age and sun exposure, for each clone, leaves were collected at different node positions from the basal to the apical parts of the shoots and at both sides of the canopies, in order to “cover” a wide range of spectral variability

within a given clone and help increasing the generalization ability of the created classifiers.

The combination of the second derivative of the normalized reflectance was highly efficient in extracting the information necessary for clone discrimination. In fact, the normalized reflectance, as well as its first derivative, seemed to contain spurious information that was "confusing" the classifiers and lowering their classification efficiency. Moreover, the small percentage of variability from the second derivative of the normalized reflectance, 39.6%, that was necessary to create highly efficient classifiers also suggests that much of the information contained in the leaf reflectance is not important to distinguish the four clones. However, there is no guarantee that the information that is not useful to discriminate the clones under analysis will not be relevant for discriminating other clones. The low dispersion of classification percentage results in the validation sets from the Monte-Carlo repetitions suggests that the created classifiers are robust.

The results obtained in the present work evidence the potential capability of local spectroscopy combined with PLS techniques to successfully discriminate among clones of a given grapevine variety. Though preliminary, these results are completely innovative and promising, and constitute the starting point of a further research involving increased number of clones per a given variety, increased number of varieties and growing sites, in order to create a wide and more complete database of classifiers. This database could be further implemented for commercial purposes, providing the grape and wine industry, including nurseries, wineries, grape growers and suppliers with an automatic, fast, environmentally friendly (solvent-free), and reliable tool for grapevine clone discrimination.



## 5. Conclusions

The present work suggests that it is possible to discriminate among grapevine leaves of different clones of Cabernet Sauvignon *(Vitis vinifera* L.) variety by using partial least squares to analyze hyperspectral imaging data taken in a narrow range of the visible and near-infrared spectrum. This narrow range, which includes strong chlorophyll absorption and the red-edge, was used to facilitate the future development of affordable dedicated spectroscopes. Although further research including more clones per variety,

increased number of grapevine varieties and sites is needed, the results obtained in this work seem to warrant the use of this new method based on leaf spectroscopy and partial least squares analysis for grapevine clone discrimination.

**Bibliography**


ALONSO-VILLAVERDE, V., BOSO, S., SANTIAGO, J.L., GAGO, P., RODRIGUEZ-GARCIA, M.I., & MARTINEZ, M.C. (2011) Leaf thickness and structure of vitis vinifera L. CV. albarino clones and its possible relation with susceptibility to downy mildew (plasmopara viticola) infection. *Journal International des Sciences de la Vigne et du Vin* **45**, 161-169

ANHALT, U. C. M., MARTINEZ, S. C., RUHL, E. & FORNECK, A. (2011). Dynamic grapevine clones-an AFLP-marker study of the *Vitis vinifera* cultivar Riesling comprising 86 clones. *Tree Genetics & Genomes* **7**, 739-746.

ARLOT, S. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys* **4,** 40-79.

ASNER, G.P., JONES, M.O., MARTIN, R.E., KNAPP, D.E., & HUGHES, R.F. (2008) Remote sensing of native and invasive species in Hawaiian forests. *Remote Sensing of Environment* **112**, 1912-1926.

BLAICH, R., KONRADI, J., RÜHL, E., & FORNECK, A. (2007) Assessing genetic variation among Pinot noir *(Vitis vinifera* L.) clones with AFLP markers. *American Journal of Enology and Viticulture* **58**, 526-529.

BURKHOLDER, A., WARNER, T.A., CULP, M., & LANDENBERGER, R. (2011) Seasonal trends in separability of leaf reflectance spectra for *Ailanthus altissima* and four other tree species. *Photogrammetric Engineering and Remote Sensing* **77,** 793-804.

CASTRO-ESAU, K.L., SÁNCHEZ-AZOFEIFA, G.A., & CAELLI, T. (2004) Discrimination of lianas and trees with leaf-level hyperspectral data. *Remote Sensing of Environment* **90**, 353-372.

CERVERA, M.T., CABEZAS, J.A., SANCHEZ-ESCRIBANO, E., CENIS, J.L., & MARTINEZ-ZAPATER, J.M. (2000) Characterization of genetic variation within table grape varieties (*Vitis vinifera* L.) based on AFLP markers. *Vitis* **39**, 109-114.

COLLINS, W. (1978) Remote sensing of crop type and maturity. *Photogrammetric Engineering and Remote Sensing* **44**, 43-55.

CONFORTO, E.C., BITTENCOURT, N.S., SCALOPPI, E.J., & BIAGI, R.M. (2011) Comparação entre folhas sombreadas de sete clones adultos de seringueira. *Revista Ceres* **58**, 29-34.

CRETAZZO, E., MENEGHETTI, S., DE ANDRÉS, M.T., GAFORIO, L., FRARE, E., & CIFRE, J. (2010) Clone differentiation and varietal identification by means of SSR, AFLP, SAMPL and M-AFLP in order to assess the clonal selection of grapevine: The case study of Manto Negro, Callet and Moll, autochthonous cultivars of Majorca. *Annals of Applied Biology* **157**, 213-227.

DOCKRAY M. (1981) *Verification of a new method for determining chlorophyll concentration in plants by remote sensing*. MSc Thesis. Centre for Environmental Technology, Imperial College, University of London. London, United Kingdom.

DURGANTE, F.M., HIGUCHI, N., ALMEIDA, A., & VICENTINI, A. (2013) Species spectral signature: Discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy. *Forest Ecology and Management* **291**, 240-248.

EHE (2006). Spectral calibration with Osram Dulux Mobil fluorescence lamp, SPECIM- Spectral Imaging Ltd. - Technical Note TN-0014.

ENGELS, J. M. M. (1983). A Systematic Description of Cacao Clones. 2. The Discriminative Value of Qualitative Characteristics and the Practical Compatability of the Discriminative Value of Quantitative and Qualitative Descriptors. *Euphytica* **32**, 387-396.

FANIZZA, G., LAMAJ, F., RESTA, P., RICCIARDI, L., & SAVINO, V. (2005) . Grapevine cvs Primitivo, Zinfandel and Crljenak kastelanski: Molecular analysis by AFLP. *Vitis* **44**, 147-148.

FERREIRO-ARMAN, M., DA COSTA, J. P., HOMAYOUNI, S., MARTIN-HERRERO, J. (2006). Hyperspectral image analysis for precision viticulture. *Lecture Notes in Computer Science* **4142**, 730-741.

FORNECK, A., WEGSCHEIDER, E., & BENJAK, A. (2009) Clonal variation in pinot noir revealed by S-SAP involving universal retrotransposon-based sequences. *American Journal of Enology and Viticulture* **60**, 104-109.

FERRANDINO, A. & GUIDONI, S. (2010). Anthocyanins, flavonols and hydroxycinnamates: an attempt to use them to discriminate *Vitis vinifera* L. cv 'Barbera' clones. *European Food Research and Technology* **230**, 417-427.

GALET, P. (1979). A practical ampelography: grapevine identification, A practical ampelography: grapevine identification, Ithaca, N.Y., Comstock Pub. Associates.

GAUSMAN, H.W., ALLEN, W.A., CARDENAS, R., & RICHARDSON, A.J. (1970) Relation of light reflectance to histological and physical evaluations of cotton leaf maturity (*Gossypium hirsutum* L.). *Applied Optics* **9**, 545-552.

GAYOL, M. F., LABUCKAS, D. O., NICOTRA, V. E., OBERTI, J. C. & GUZMAN, C. A. (2009). Simmondsins quantification by spectrophotometry UV-vis in press cake and in jojoba seeds (*Simmondsia chinensis* (Link) Schneider), from Argentina. *Industrial Crops and Products* **29**, 177-181.

GELADI, P. & KOWALSKI, B. R. (1986). Partial Least-Squares Regression - a Tutorial. *Analytica Chimica Acta* **185**, 1-17.

GRANT, L. (1987). Diffuse and specular characteristics of leaf reflectance. *Remote Sensing of Environment* **22**, 309-322.

IMAZIO, S., LABRA, M., GRASSI, F., WINFIELD, M., BARDINI, M. & SCIENZA, A. (2002). Molecular tools for clone identification: the case of the grapevine cultivar 'Traminer'. *Plant Breeding* **121**, 531-535.

KONRADI J., BLAICH R., & FORNECK A. (2007) Genetic variation among clones and sports of ‘‘Pinot noir’’ (*Vitis vinifera* L.). *European Journal of Horticultural Science* **72**, 275–279.

KOZJAK P., KOROˇSEC-KORUZA Z., & JAVORNIK B. (2003) Characterization of cv. Refoˇsk (*Vitis vinifera* L.) by SSR markers. *Vitis* **42**, 83–86.

LACAR, F. M., LEWIS, M. M., GRIERSON, I. T., (2001). Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia. *Geoscience and Remote Sensing Symposium* (*2001-IGARSS '01*), pp. 2875-2877. 9-13 July 2001, Sydney, Australia.

LIN, D.-D., ZHANG, G.-F., YU, J.-B., & FENG, J. (2011) Analyses of photosynthetic pigment content and chlorophyll fluorescence parameter in leaves of different

clones of Cinnamomum camphora. *Journal of Plant Resources and Environment* **20**, 56-61.

MONCADA, X. & HINRICHSEN, P. (2007) Limited genetic diversity among clones of red wine cultivar 'Carmenère' as revealed by microsatellite and AFLP markers. *Vitis* **46**, 174-180.

OSMAN, F. N., HASHIM, H., AL-JUNID, S. A. M., HARON, M. A. B., ABDULLAH, N. E. & MUHAMMAD, M. F. B. (2010). A Statistical Approach for Rubber Seed Clones Classification Using Reflectance Index. *Proceedings of the IV Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation (AMS),*pp. 291-295. 26- 28 May 2010, Kota Kinabalu, Borneo.

PANDE, P. (2011). Variation in Wood Properties and Growth in Some Clones of Populus deltoides Bartr. ex Marsh. *American Journal of Plant Sciences* **2**, 644-649.

PENUELAS, J., FILELLA, I., BIEL, C., SERRANO, L. & SAVE, R. (1993). The reflectance at the 950-970 nm region as an indicator of plant water status. *International Journal of Remote Sensing* **14**, 1887-1905.

REGNER, F., WIEDECK, E. & STADLBAUER, A. (2000). Differentiation and identification of White Riesling clones by genetic markers. *Vitis* **39**, 103-107.

REGNER F., HACK R., & SANTIAGO J.L. (2006) Highly variable *Vitis* microsatellite loci for identification of Pinot noir clones. *Vitis*, **45** 85–91

RIBEIRO DA LUZ, B. (2006) Attenuated total reflectance spectroscopy of plant leaves: a tool for ecological and botanical studies. *New Phytologist* **172**, 305-318.

SCHLAPFER, F. & FISCHER, M. (1998). An isozyme study of clone diversity and relative importance of sexual and vegetative recruitment in the grass Brachypodium pinnatum. *Ecography* **21**, 351-360.

SÁNCHEZ-AZOFEIFA, G.A., CASTRO, K., WRIGHT, S.J., GAMON, J., KALACSKA, M., RIVARD, B., SCHNITZER, S.A. & FENG, J.L. (2009) Differences in leaf traits, leaf internal structure, and spectral reflectance between two communities of lianas and trees: Implications for remote sensing in tropical environments. *Remote Sensing of Environment* **113**, 2076-2088.

SILVESTRONI O., DI PIETRO D., INTRIERI C., VIGNANI R., FILIPPETTI I., DEL CASINO C., SCALI M., & CRESTI M. (1997) Detection of genetic diversity among clones of cv. Fontana (*Vitis vinifera* L.) by microsatellite DNA polymorphism analysis. *Vitis* **36**, 147–150.

SLATON, M. R., HUNT, E. R., & SMITH, W. K. (2001). Estimating near-infrared leaf reflectance from leaf structural characteristics. *American Journal of Botany* **88**, 278-284

TAYLOR, S., BAKER, D., OWUOR, P., ORCHARD, J., OTHIENO, C., & GAY, C. (1992) A model for predicting black tea quality from the carotenoid and chlorophyll composition of fresh green tea leaf. *Journal of the Science of Food and Agriculture* **58**, 185-191.

TERASHIMA, I., & SAEKI, T. (1983) Light environment within a leaf I. Optical properties of paradermal sections of camellia leaves with special reference to differences in the optical properties of palisade and spongy tissues. *Plant and Cell Physiology* **24**, 1493-1501

VOGELMANN, T. C. (1993). Plant-Tissue Optics. *Annual Review of Plant Physiology and Plant Molecular Biology* **44**, 231-251

Figure 1- Experimental setup for hyperspectral imaging of leaves.

Figure 2- Envelope (minimum and maximum value) curves containing: a) the normalized reflectance; b) the first derivative of the normalized reflectance; and, c) the second derivative of the normalized reflectance. The curves may have the same color for different clones because the values were similar.

Figure 3- Correctly classified leaf percentage for the validation set in the 100 repetitions of Monte-Carlo cross-validation method for the four clones to be discriminated. The classifiers may have three inputs: a) the normalized reflectance spectra, prefix "r"; b) the first derivative of the normalized reflectance, prefix "d1"; and, c) the second derivative of the normalized reflectance "d2". The box represents the $25^{th}$, $50^{th}$ and $75^{th}$ percentiles, while the whiskers represent the $5^{th}$ and $95^{th}$ percentiles. The crosses stand for the $1^{st}$ and $99^{th}$ percentiles and the minimum and maximum values. When the $1^{st}$ or $99^{th}$ percentiles coincide with the minimum and maximum values, respectively, only one cross is visible. The dark square represents the mean value. The arrows point to the median line.

Figure 4- Box plot of the Root Mean Squared Error (RMSE) for the validation set in the 100 repetitions of Monte-Carlo cross-validation method. These results are relative to the four output variables and corresponding clones of the classifiers whose input was the second derivative of the normalized reflectance. The box represents the $25^{th}$, $50^{th}$ and $75^{th}$ percentiles, while the whiskers represent the $5^{th}$ and $95^{th}$ percentiles. The crosses stand for the $1^{st}$ and $99^{th}$ percentiles and the minimum and maximum values. When the $1^{st}$ or $99^{th}$ percentiles coincide with the minimum and maximum values, respectively, only one cross is visible. The dark square represents the mean value.

Figure 5 - Variance of the second derivative of reflectance and of the clone class, in the whole sample set, that is explained by the PLS components.

Figure 6- Histogram of the best number of components in the 100 Monte-Carlo cross-validation repetitions for classifiers whose input was the second derivative of reflectance.

Figure 7– Loadings of the PLS components for the whole sample set when classifier input was the second derivative of the normalized reflectance. Component rank: a) first; b) second; c) third; d) fourth; e) fifth; f) sixth; g) seventh; h) eighth; i) ninth.

Table 1. Designation of Cabernet Sauvignon *(Vitis vinifera* L.) clones used in this study and information about their origin and selection.

| Feature | Cabernet Sauvignon clones | | | |
|---|---|---|---|---|
| | **15** | **169** | **685** | **R5** |
| Selection | ENTAV 1 | ENTAV 2 | ENTAV 157 | VCR R5 |
| Place of origin | Gironde (France) | Gironde (France) | Atlantic Pyrenees (France) | San Michelle all'Adige (Italy) |
| Agreement Yr | 1971 | 1972 | 1980 | 1988 |

Table 2 – Mean percentage of correctly classified leaves in the validation sets of the Monte-Carlo cross- validation method.

| | CS 15 | CS 169 | CS 685 | CS R5 |
|---|---|---|---|---|
| Normalized reflectance | 94.4 | 92.4 | 91.4 | 94.2 |
| Normalized reflectance 1st derivative | 97.4 | 95.8 | 97.8 | 99.2 |
| Normalized reflectance 2nd derivative | 98.2 | 99.2 | 100 | 97.8 |

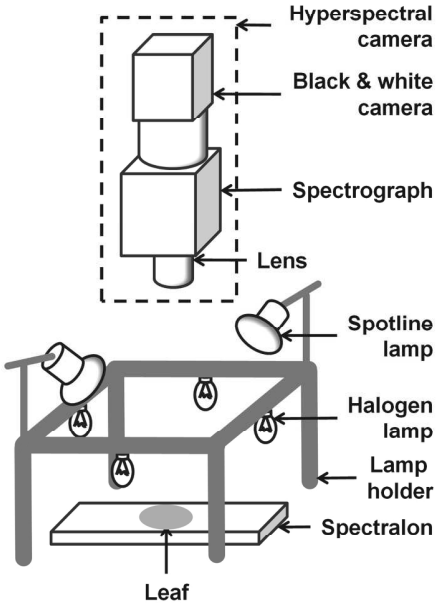


Figure 1- Experimental setup for hyperspectral imaging of leaves.
190x142mm (300 x 300 DPI)

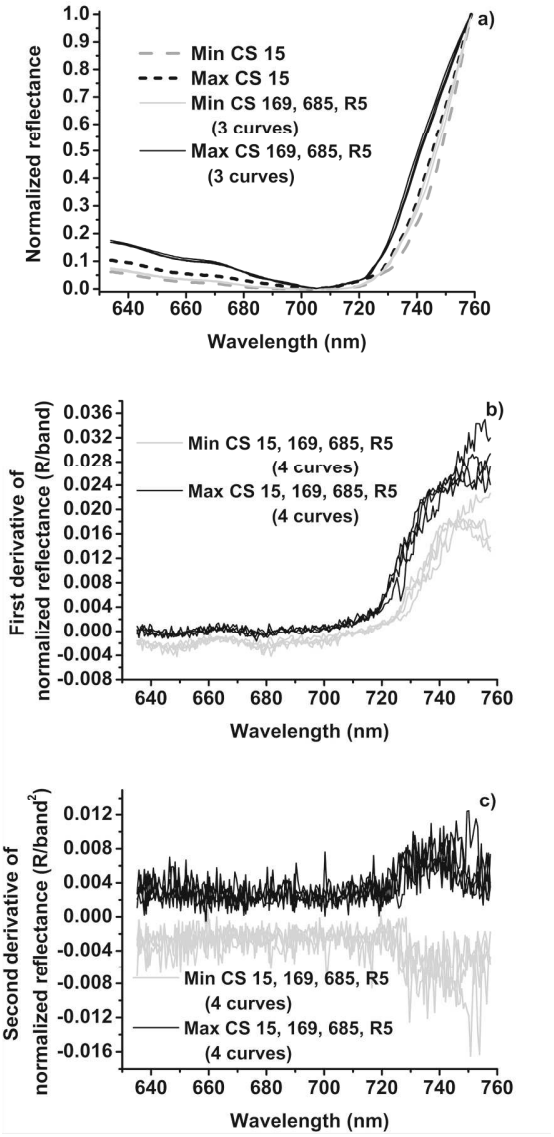


Figure 2- Envelope (minimum and maximum value) curves containing: a) the normalized reflectance; b) the first derivative of the normalized reflectance; and, c) the second derivative of the normalized reflectance. The curves may have the same color for different clones because the values were similar.
188x400mm (200 x 200 DPI)

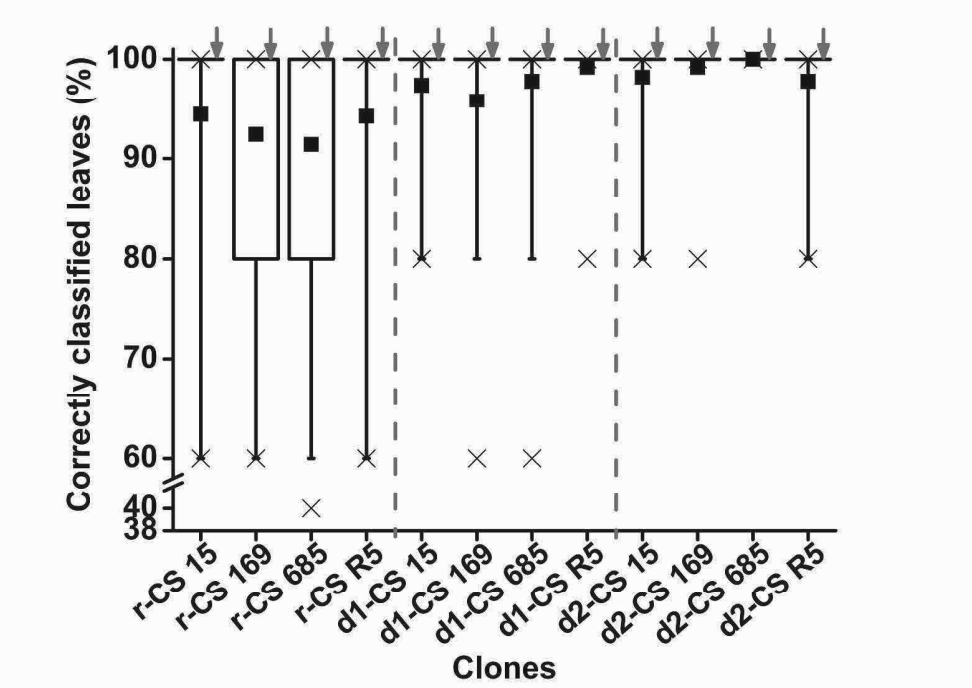


Figure 3- Correctly classified leaf percentage for the validation set in the 100 repetitions of Monte-Carlo cross-validation method for the four clones to be discriminated. The classifiers may have three inputs: a) the normalized reflectance spectra, prefix "r"; b) the first derivative of the normalized reflectance, prefix "d1"; and, c) the second derivative of the normalized reflectance "d2". The box represents the 25th, 50th and 75th percentiles, while the whiskers represent the 5th and 95th percentiles. The crosses stand for the 1st and 99th percentiles and the minimum and maximum values. When the 1st or 99th percentiles coincide with the minimum and maximum values, respectively, only one cross is visible. The dark square represents the mean value. The arrows point to the median line.
270x192mm (300 x 300 DPI)

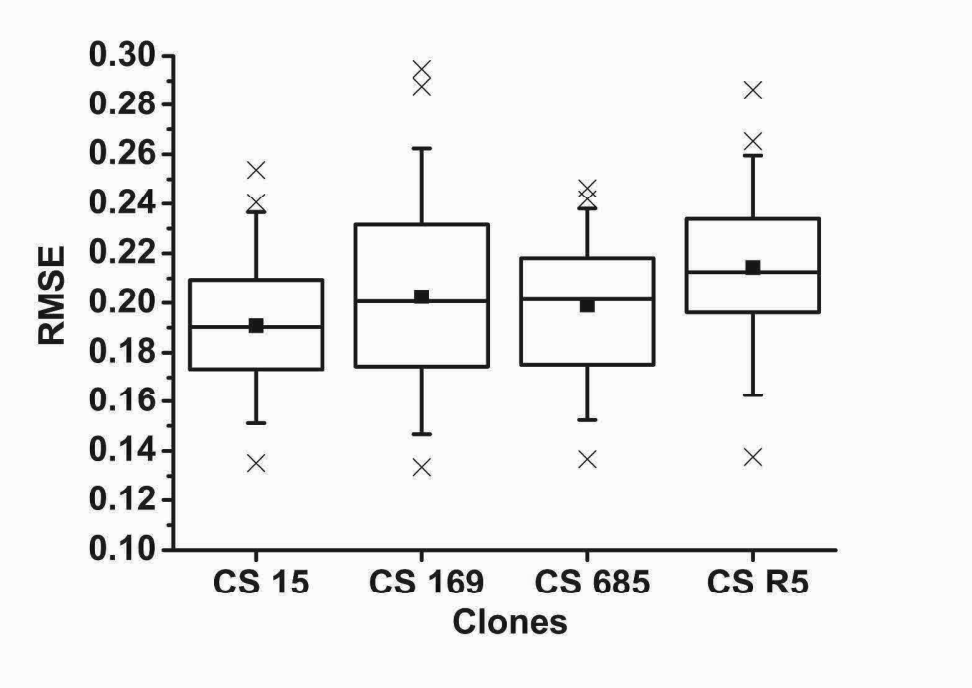


Figure 4- Box plot of the Root Mean Squared Error (RMSE) for the validation set in the 100 repetitions of Monte-Carlo cross-validation method. These results are relative to the four output variables and corresponding clones of the classifiers whose input was the second derivative of the normalized reflectance. The box represents the 25th, 50th and 75th percentiles, while the whiskers represent the 5th and 95th percentiles. The crosses stand for the 1st and 99th percentiles and the minimum and maximum values. When the 1st or 99th percentiles coincide with the minimum and maximum values, respectively, only one cross is visible. The dark square represents the mean value.
270x192mm (300 x 300 DPI)

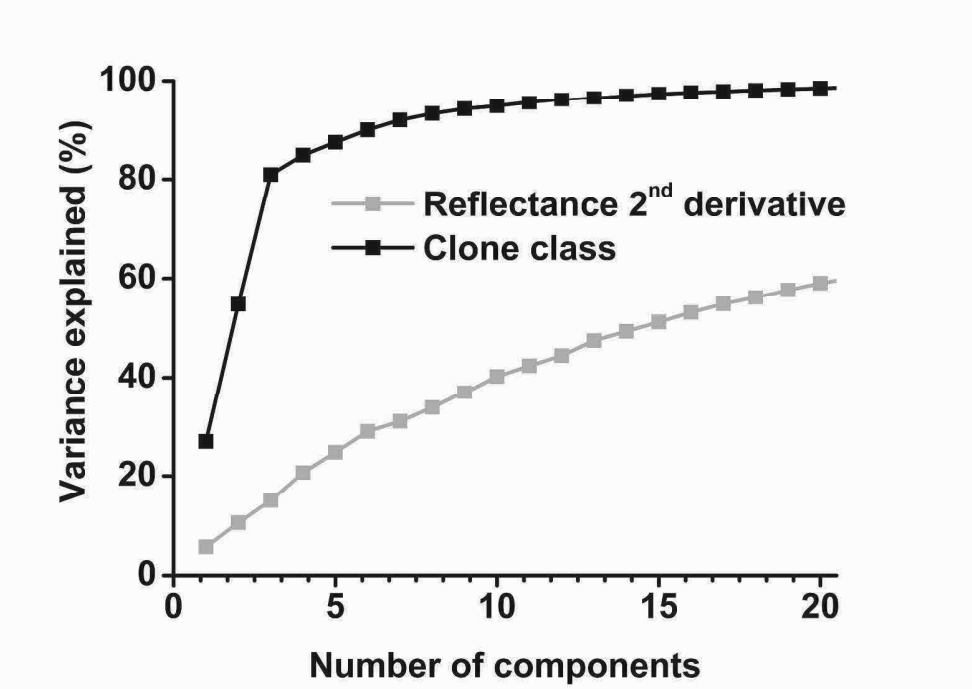


Figure 5 - Variance of the second derivative of reflectance and of the clone class, in the whole sample set, that is explained by the PLS components.
270x192mm (300 x 300 DPI)

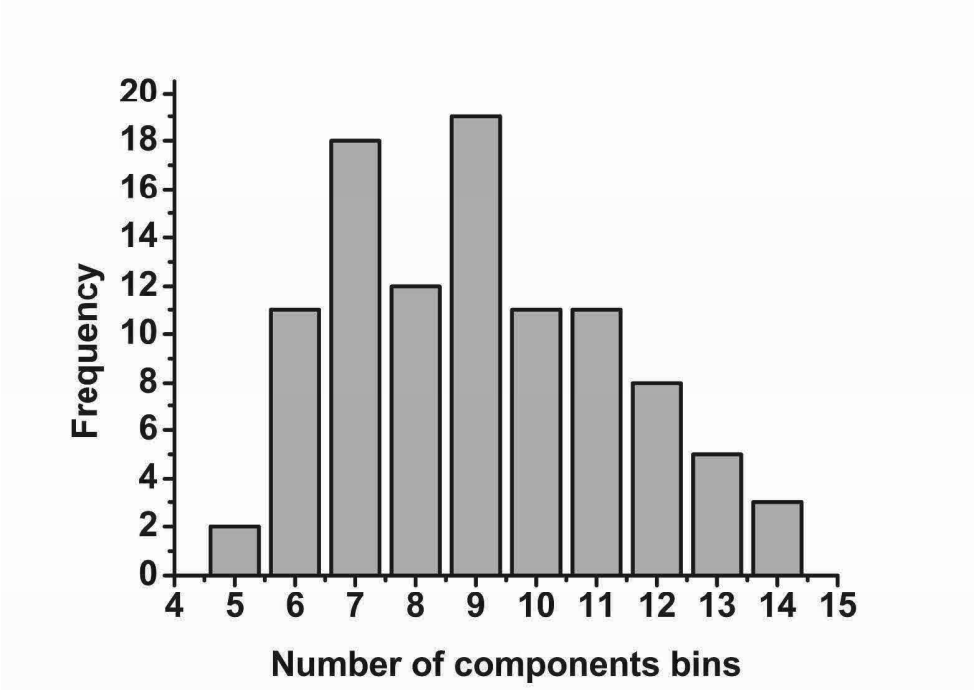


Figure 6- Histogram of the best number of components in the 100 Monte-Carlo cross-validation repetitions for classifiers whose input was the second derivative of reflectance.
270x192mm (300 x 300 DPI)

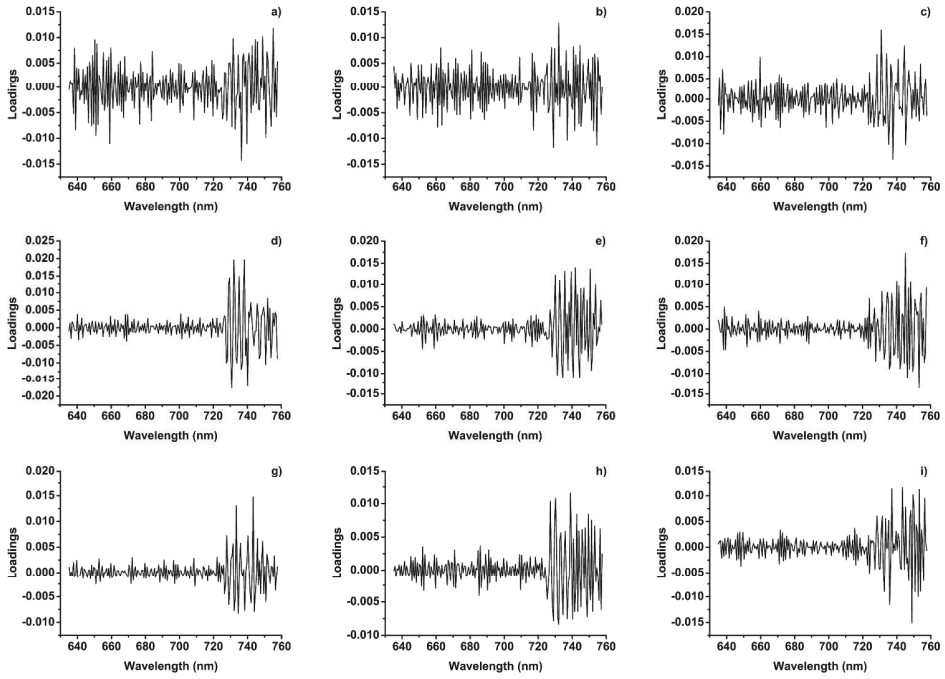


Figure 7– Loadings of the PLS components for the whole sample set when classifier input was the second derivative of the normalized reflectance. Component rank: a) first; b) second; c) third; d) fourth; e) fifth; f) sixth; g) seventh; h) eighth; i) ninth.
284x201mm (300 x 300 DPI)