



Data Mining and Non-Invasive Proximal Sensing for Precision Viticulture

Salvador Gutiérrez, Javier Tardaguila, Juan Fernández-Novales and Maria P. Diago*

¹ Instituto de Ciencias de la Vid y del Vino (University of La Rioja, CSIC, Gobierno de La Rioja) Ctra. De Burgos Km. 6, 26007. Logroño, La Rioja. Spain; E-Mails: salvador.gutierrez@unirioja.es, javier.tardaguila@unirioja.es, juan.fernandezn@unirioja.es, mpaz.diago.santamaria@gmail.com

* Author to whom correspondence should be addressed; E-Mail: mpaz.diago.santamaria@gmail.com; Tel.: +34941894980-ext.410065; Fax: +34941899728.

Published: 12 November 2015

Abstract: Modern and sustainable viticulture entails objective and fast monitoring of crucial variables for rational decision making. Data mining strategies may be applied to agricultural data, with the aim of yielding useful, reliable and objective information. This work presents the most recent applications of machine learning algorithms to grapevine plant phenotyping, such as varietal discrimination and water status assessment. Support vector machine (SVM) and modified partial least squares (MPLS) models were built using NIR spectra acquired in the vineyard, on grapevine leaves, with a portable spectrophotometer working on the spectral range between 1600 to 2500 nm. Spectral measurements were acquired on the adaxial side of 200 individual leaves (20 leaves per cultivar) of ten (*Vitis vinifera* L.) varieties. Sequential minimal optimization (SMO) algorithm was used for the training of a SVM for varietal discrimination. The classifier's performance for the 10 varieties surpassed the 94.9% mark. For water status assessment, the predictive model based on MPLS using the reflectance spectra of four cultivars, and the first and second derivative, yielded a $R^2 = 0.83$ for stem water potential (Ψ_{stem}), which is widely recognized as an integrative indicator of whole-vine water status, but destructive and very laborious. These results show the power of the combined use of data mining and non-invasive sensing for grapevine phenotyping and their usefulness for the wine industry.

Keywords: machine learning; grapevine phenotyping; infrared spectra; grapevine cultivar; proximal sensing; water stress; stem water potential

1. Introduction

The study of grapevine phenotype, which involves its physical and biochemical traits as a result of the interaction of its genotype and the environment, is a key topic in modern viticulture. There are thousands of *Vitis vinifera* L. grapevine varieties worldwide [1] and their cultivation, wine quality potential and the price paid for their grapes is variety dependent [2]. Also, the ability of set up irrigation rules is based on soil water balance calculations or direct measures of soil moisture. Nonetheless, these methods could not be totally reliable due to soil heterogeneity and are likely to return cumulative errors [3].

Current methods to address variety discrimination include visual ampelometry [1], wet chemistry genetic [4] and isoenzyme analyses [5], and very recently hyperspectral imaging under laboratory conditions [6]. However, identification of grapevine varieties under field conditions in order to recognize grapevines of not-allowed cultivars in certain appellation regions worldwide or unknown vines in older vineyards, where more than a single cultivar was often planted is of great interest in viticulture.

In terms of the plant's water status, midday stem water potential (Ψ_{stem}) has been widely accepted as a useful and reproducible index of the plant water status, also proposed as a more integrative indicator of vine water condition [7], yet still this method requires time-consuming and destructive ways to carry out the measurement.

Spectroscopy is based on the interaction of electromagnetic radiation with matter at different wavelengths. The use of local spectroscopy for plant variety discrimination has been applied in controlled indoor environments in several crops, such as bayberry [8] and strawberry [9] using LOCAL algorithm and partial least squares as discrimination methods, respectively, and outdoors in tomato [10]. Several works have demonstrated the suitability of leaf reflectance to evaluate grapevine and other crops water status [12–15].

This work discusses the use of data mining applied to non-invasively retrieved data to yield key information for plant phenotyping in the frame of precision viticulture.

2. Materials and methods

2.1. Spectra collection

For varietal discrimination, spectral measurements were acquired under field conditions on the adaxial side of 200 individual leaves (20 leaves per cultivar) of 5 red: Cabernet Sauvignon, Carmenere, Tempranillo, Pinot Noir, Caladoc, and 5 white: Viura, Treixadura, White Grenache, Pedro Ximenez, Viognier (*Vitis vinifera* L.) varieties. For stem water potential estimation, spectral measurements were acquired under field conditions on the adaxial side of 80 individual leaves (20 leaves per cultivar) of 4 red: Tempranillo, Grenache, Cabernet Sauvignon and Marselan (*Vitis vinifera* L.) varieties. The stem water potential was measured after spectra acquisition using a Scholander pressure Chamber (Model 600, PMS Instruments Co., Albany, USA). All measurements were taken in a commercial vineyard located in Navarra (Spain) during the ripening period of season 2012.

An integrated handheld NIR spectral analyzer (microPHAZIR™, Thermo Fisher Scientific Inc., Waltham, MA, USA), working in reflectance mode (log 1/R) in the range of 1600–2400 nm with a non-constant interval of 8.7 nm (pixel resolution 8 nm, optical resolution 12 nm) was used.

Sensor integration time was 600 ms. The device was equipped with quartz protection to prevent dirt accumulation. For each leaf, five spectral measurements were acquired and sample temperature during measurements ranged between 23 to 25°C. Vinyl gloves were used at all times to handle the leaves to not distort them with external pollutants from hand manipulation.

Figure 1 shows a block diagram describing the experimental setup.

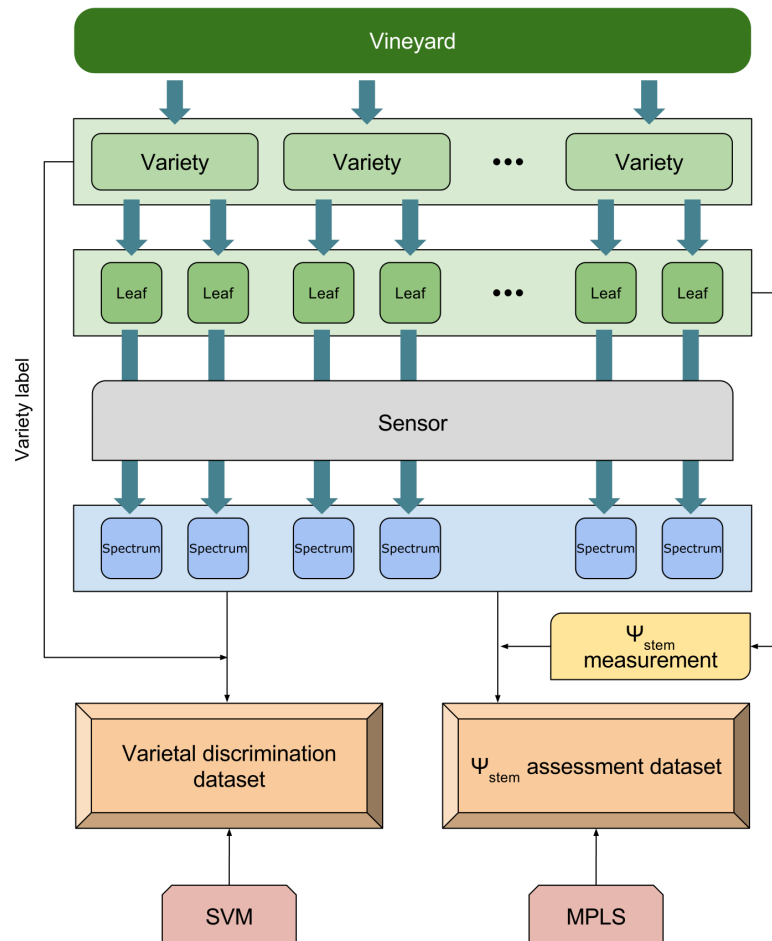


Figure 1. Experimental setup.

2.2. Spectra processing for variety discrimination

Due to measurement spectral inaccuracies, one sample from Treixadura and another one from White Grenache were removed. For all cultivars, the average from the five spectral measurements per leaf was computed and considered the average spectrum per leaf. In order to avoid scattering issues, several pre-processing algorithms were applied to the spectra: Standard Normal Variates (SNV) followed by a detrending and a first grade Savitzky-Golay filter (window size: 5).

Support vector machines (SVMs) were used as a learning method for the training of the variety discrimination model. A SVM was trained using the Sequential Minimal Optimisation (SMO) algorithm [16] with polynomial kernel and a complex parameter C equals to 3.5.

The accuracy of the trained model was evaluated by using cross-validation and the percentage of correctly-classified examples.

2.3. Spectra processing for water stress estimation

For all cultivars, the average from the five spectral measurements per leaf was computed and considered the average spectrum per leaf. Principal component analysis (PCA) was used to reduce the dimensionality of the data to a smaller number of components, to examine any possible grouping and to visualize the presence of outliers based on Global Mahalanobis (GH) distance [17].

As spectral pre-treatments, the Standard Normal Variate (SNV) plus Detrending (DT) procedure was used to remove the multiplicative interferences of scatter, in addition to a derivative mathematical treatment. Modified Partial Least Squares (MPLS) regression was tested for the prediction of Ψ_{stem} . To prevent over-fitting, the assessment of the calibration model was performed by cross-validation. Chemometric analysis was performed using the WinISI II software package version 1.50 (Infrasoft International, Port Matilda, PA, USA) and the Unscrambler software package version 9.1 (CAMO ASA, Oslo, Norway).

3. Results and Discussion

3.1. Varietal discrimination

The results obtained with the Sequential Minimal Optimization algorithm for the discrimination of 10 grapevine varieties are shown in Table 1.

Table 1. Confusion matrix of the classifier. V: Viura, T: Treixadura, WG: White Grenache, PX: Pedro Ximenez, Vi: Viognier, CS: Cabernet Sauvignon, C: Carmenere, Te: Tempranillo, PN: Pinot Noir, Ca: Caladoc. %: Variety correctly classified percentage.

	Classified as										%
	V	T	WG	PX	Vi	CS	C	TE	PN	Ca	
V	20	0	0	0	0	0	0	0	0	0	100
T	0	17	0	0	1	0	0	0	1	0	89.5
WG	0	0	19	0	0	0	0	0	0	0	100
PX	1	0	0	18	1	0	0	0	0	0	90
Vi	0	0	1	0	18	0	0	0	0	1	90
CS	0	0	0	0	0	19	1	0	0	0	95
C	0	0	0	0	0	1	19	0	0	0	95
Te	0	0	1	0	0	0	0	19	0	0	95
PN	0	0	0	0	0	0	0	0	20	0	100
Ca	0	0	0	0	1	0	0	0	0	19	95

The global output resulted in a 95% accuracy in the discrimination of the whole set of varieties, where 188 out of 198 instances were correctly classified.

From an individual point of view, 3 varieties (Viura, White Grenache, Pinot Noir) have reached a perfect score (100% of correct discrimination), while the remaining varieties obtained values between 90% and 95%.

As it can be seen from the confusion matrix in Table 1, two samples of Pedro Ximenez were misclassified into Viognier and White Grenache, both being white varieties. The same applied to red varieties, such as Cabernet Sauvignon and Carmenere, where both exhibited a misclassified sample. One sample from a white variety (Viognier) was incorrectly classified as a red one (Caladoc). Finally, two red samples, one from Tempranillo and another one from Caladoc, were misclassified as white varieties, White Grenache and Viognier, respectively.

NIR information has been used for cultivar identification in plum [20] and strawberry [9] using the spectra acquired on the fruits. Also, leaves have been recently used for grapevine discrimination through hyperspectral imaging [6] under laboratory conditions and using the visible range of 380 to 1028 nm.

3.2. Assessment of plant water status

Table 2 shows the best outcomes obtained for Ψ_{stem} assessment. The correlation coefficient reached the 0.83 mark in cross-validation, exposing a higher accuracy than that reported in a recent work [15] (SECV = 0.18 and $R^2 = 0.71$).

Table 2. Calibration and cross-validation results for stem water potential (Ψ_{stem}).

Calibration		Cross Validation	
SEC	R^2_C	SECV	R^2_{CV}
0.13	0.86	0.14	0.83

These results present discrimination and regression models with a high precision, that shows how accurate the information retrieved from non-invasively acquired NIR spectra in the range of 1600 to 2500 nm of grapevine leaves can be for the discrimination of grapevine varieties and the assessment of vineyard water status using different data mining approaches.

4. Conclusions

The remarkable performance of the developed models under field conditions paves the way for the use of data mining algorithms in combination with non-invasive sensing tools – such as a portable NIR analyzer – as powerful phenotyping and water stress assessment methods in viticulture and other crops. These results open a gateway for the fast and non-destructive varietal classification and water status estimation in viticulture and – potentially – in general agriculture.

Acknowledgments

This work has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement N 311775 (KBBE-INNOVINE).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Galet, P.; others. *A practical ampelography*; Cornell University Press., 1979.
2. Clarke, O.; Rand, M. *Grapes & Wines: A comprehensive guide to varieties and flavours*; Pavilion Books, 2015.
3. Weckler, P.R.; Maness, N.O.; Stone, M.L. Estimating Water Stress in Plants Using Hyperspectral Sensing. *system* **2004**, *1*, 4.
4. Pelsy, F.; Hocquigny, S.; Moncada, X.; Barbeau, G.; Forget, D.; Hinrichsen, P.; Merdinoglu, D. An extensive study of the genetic diversity within seven French wine grape variety collections. *Theoretical and Applied Genetics* **2010**, *120*, 1219–1231.
5. Altube, H.; CABELLO, F.; ORTIZ, J.M. Caracterización de variedades y portainjertos de vid mediante isoenzimas de los sarmientos. *Vitis* **1991**, *30*, 203–212.
6. Diago, M.P.; Fernandes, A.; Millan, B.; Tardaguila, J.; Melo-Pinto, P. Identification of grapevine varieties using leaf spectroscopy and partial least squares. *Computers and electronics in agriculture* **2013**, *99*, 7–13.
7. Choné, X.; Van Leeuwen, C.; Dubourdieu, D.; Gaudillère, J.P. Stem water potential is a sensitive indicator of grapevine water status. *Annals of botany* **2001**, *87*, 477–483.
8. Li, X.; He, Y.; Fang, H. Non-destructive discrimination of Chinese bayberry varieties using Vis/NIR spectroscopy. *Journal of Food Engineering* **2007**, *81*, 357–363.
9. Sánchez, M.T.; De la Haba, M.J.; Benítez-López, M.; Fernández-Navales, J.; Garrido-Varo, A.; Pérez-Marín, D. Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *Journal of Food engineering* **2012**, *110*, 102–108.
10. Xu, H.r.; Yu, P.; Fu, X.p.; Ying, Y.b. On-site variety discrimination of tomato plant using visible-near infrared reflectance spectroscopy. *Journal of Zhejiang University Science B* **2009**, *10*, 126–132.
11. Chauchard, F.; Cogdill, R.; Roussel, S.; Roger, J.; Bellon-Maurel, V. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems* **2004**, *71*, 141–150.
12. Gaulton, R.; Danson, F.; Ramirez, F.; Gunawan, O. The potential of dual-wavelength laser scanning for estimating vegetation moisture content. *Remote Sensing of Environment* **2013**, *132*, 32–39.
13. Santos, A.O.; Kaye, O. Grapevine leaf water potential based upon near infrared spectroscopy. *Scientia Agricola* **2009**, *66*, 287–292.
14. Suárez, L.; Zarco-Tejada, P.J.; González-Dugo, V.; Berni, J.; Sagardoy, R.; Morales, F.; Fereres, E. Detecting water stress effects on fruit quality in orchards with time-series PRI airborne imagery. *Remote Sensing of Environment* **2010**, *114*, 286–298.
15. De Bei, R.; Cozzolino, D.; Sullivan, W.; Cynkar, W.; Fuentes, S.; Damberg, R.; Pech, J.; Tyerman, S. Non-destructive measurement of grapevine water potential using near infrared spectroscopy. *Australian Journal of Grape and Wine Research* **2011**, *17*, 62–71.
16. Platt, J.C.; others. Using analytic QP and sparseness to speed training of support vector machines. *Advances in neural information processing systems* **1999**, pp. 557–563.

17. Shenk, J.S.; Westerhaus, M.O. Calibration the ISI way. *Near infrared spectroscopy: The future waves* **1996**, pp. 198–202.
18. Fu, X.; Zhou, Y.; Ying, Y.; Lu, H.; Xu, H. Discrimination of pear varieties using three classification methods based on near-infrared spectroscopy. *Transactions of the ASABE* **2007**, *50*, 1355–1361.
19. Fernández-Navales, J.; López, M.I.; Sánchez, M.T.; García-Mesa, J.A.; González-Caballero, V. Assessment of quality parameters in grapes during ripening using a miniature fiber-optic near-infrared spectrometer. *International journal of food sciences and nutrition* **2009**, *60*, 265–277.
20. Pérez-Marín, D.; Sánchez, M.T.; Paz, P.; Soriano, M.A.; Guerrero, J.E.; Garrido-Varo, A. Non-destructive determination of quality parameters in nectarines during on-tree ripening and postharvest storage. *Postharvest biology and technology* **2009**, *52*, 180–188.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).