Vitis vinifera phenotyping by NIR spectroscopy and chemometrics

Basile T. [1]*, *Forleo L.R.* [1], *Cardone M.F.* [1], *Bergamini C.* [1]

[1] CREA Research Centre for Viticulture and Enology, Via Casamassima 70010 Turi (BA), Italy
*e-mail: teodora.basile@crea.gov.it

State-of-the-art: Plant phenotyping is usually performed with sampling methods that are costly, labor-intensive, time-consuming, and destructive. In modern plant phenotyping, several non-destructive techniques have been developed with numerous advantages over the traditional ones. In this study, a multivariate analysis combined with near-infrared (NIR) spectroscopy was employed to classify grapevine leaves.

Grape varieties: Red Globe and Sugraone

Sampling: Leaves of various ages (young, intermediate, and mature) were collected from vines of equal age, grown in the same vineyard (Apulia region, Southern Italy), and subjected to the same field treatments, ensuring identical *terroir*. An acclimatization step (stems immersed in water) was performed to compensate for different solar exposure. Six points on each leaf face were measured. The spectra of the lower faces were used in this work.
Instrument: TANGO FT-NIR spectrophotometer (Bruker, Germany). Data acquisition software: OPUS/QUANT version 2.0 (Bruker Optik GmbH, Ettlingen, Germany). Parameters: 12,000–4000 cm^{-1} (833–2500 nm), 8 cm^{-1} resolution and 64 scans. Statistical Analysis Software: R version 4.1.2 (2021-11-01). The R packages: caret, MASS, and mdatools.

Figure 1. TANGO NIR spectrophotometer

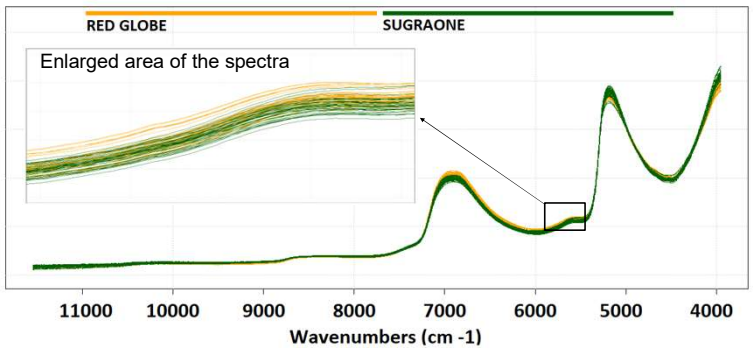

Figure 2. SNV pre-treated spectra

Development of supervised classification models

The classification techniques compared were: Linear Discriminant Analysis (LDA), Classification and Regression Trees, k-Nearest Neighbors (K-NN), Support Vector Machines (SNV), and Random Forest. A class balanced random 70/30 split of the data set was carried out. The validation of the training set was performed with a 5-fold cross-validation repeated 10 times. The best-fit model on the test data set was obtained with a K-NN based on a principal component analysis (PCA) selection (1066 wave numbers) of Standard Normal Variate (SNV) pre-treated NIR spectral data (Table 1). Anyway, the best prediction on an external test set was obtained with an LDA model (Table 2).

Table 1. Classification models with the PCA selected wave numbers (training set)

Accuracy		Kappa	
lda	0.8285714	lda	0.6447273
cart	0.8119048	cart	0.6165955
knn	0.8738095	knn	0.7444638
svm	0.8119048	svm	0.6169275
rf	0.8285714	rf	0.6502609

Table 2. Confusion matrix on the test set for the LDA model

Accuracy : 0.9231 Kappa : 0.8434		Reference	
		Red Globe	Sugraone
Prediction	Red Globe	14	2
	Sugraone	0	10

CONCLUSIONS: The discrimination ability of the NIR technique could thus be used as a tool for fast variety recognition. In the ongoing analysis of different grape varieties growing in our experimental vineyards, we plan to improve the discriminating capacity of the technique.This work shows an example of how the application of chemometric methods could be effectively used to support the decision-making process in precision agriculture or digital farming.