

Total soluble solids in grape must estimation using VIS-NIR-SWIR reflectance measured in fresh berries

Karen Brigitte Mejía-Correal ^{1*}, Víctor Marcelo ², Enoc Sanz-Ablanedo ¹ and José Ramón Rodríguez-Pérez ¹

¹ Grupo de Investigación en Geomática e Ingeniería Cartográfica (GEOINCA), Universidad de León, Avenida de Astorga sn, 24401 Ponferrada, León, Spain; kmej@unileon.es (K.B.M.-C.); esana@unileon.es (E.S.-A.); jr.rodriguez@unileon.es (J.R.R.-P.)

² Departamento de Ingeniería y Ciencias Agrarias, Universidad de León, Avenida de Astorga sn, 24401 Ponferrada, León, Spain; v.marcelo@unileon.es (V. M.)

* Correspondence: kmej@unileon.es

Supplementary Material S1

In order to support the use of PLS regression models in this research, detailed information about Random Forest algorithm tested in a previous work is included. This work consisted of a comparative analysis between Random Forest and Partial Least Square regression models for estimating TSS content in grapes (the same grape samples using in this paper). The regression models were developed R Statistical Software (version 4.2.0; R Core Team 2022) and RStudio (version 2022.07.2+576; RStudio Team, 2022). The packages used were: radomForest (v4.7.1.1; Liaw and Wiener, 2002), pls (v2.8.1; Liland et al., 2022) and caret (v6.0.93; Kuhn, 2022).

In order to identify the most suitable wavelength range to estimate TSS, 4 spectral subsets were stabilized: VIS (400-700 nm), NIR (701-1000 nm), SWIR (1001-2500 nm) and full range (400-2500 nm). The regression models were evaluated using LOOCV, R^2 , RMSE, RPD and number of factors in PLS models. As a result, we obtained in general that the PLS regression models were much better than RF regression models in order to estimate TSS in grapes. Table A1 shows the results achieved with RF models in the 4 grape varieties and different spectral range (VIS, NIR, SWIR and full range). The best result of each variety is in bold.

For white grapes the best results were obtained with raw data in NIR range (RMSE=1.64; R^2 =0.40 and RPD=2.23) for Godello and for Verdejo with SNV-transformation data in SWIR range (RMSE=1.80; R^2 =0.34 and RPD=1.87). On the other hand, red grapes achieved the best results with SNV-transformation data in NIR range (RMSE=0.81; R^2 =0.59 and RPD=2.53) for Mencía and in full range (RMSE=1.26; R^2 =0.47 and RPD=2.85) for Tempranillo. The grape variety with the best validation results is Mencía with a RMSE less than 1°Brix (RMSE=0.81°Brix), R^2 highest than 0.50 (R^2 =0.59) and RPD=2.85 which shows the accuracy of the prediction model.

Owing to RF estimation models was not as good as PLS models, and take into account the results (Table A1), we decided not using this comparative analysis in this paper.

Table S1. Cross-validation statistics for Random Forest regression models using VIS (400-700 nm), NIR (701-1000 nm), SWIR (1001-2500 nm), and full spectral range (VIS+NIR+SWIR: 400-2500 nm) for raw and SNV-transformed data.

Varieties	Pre-processed	Spectral range (nm)	RMSE (°Brix)	R^2	RPD
Godello	Raw	400-2500	1.87	0.23	2.08
		400-700*	1.64	0.40	2.23
		701-1000	1.92	0.23	2.00

Varieties	Pre-processed	Spectral range (nm)	RMSE (°Brix)	R ²	RPD
Verdejo	SNV	1001-2500	2.15	0.18	1.82
		400-2500	1.88	0.22	2.07
		400-700	1.70	0.34	2.17
		701-1000	1.89	0.24	2.06
		1001-2500	1.97	0.18	1.77
	Raw	400-2500	1.81	0.23	2.03
		400-700	1.88	0.22	2.16
		701-1000	1.83	0.26	1.98
		1001-2500	1.91	0.24	1.86
	SNV	400-2500	1.78	0.30	1.96
		400-700	1.85	0.25	2.15
		701-1000*	1.80	0.34	1.87
		1001-2500	1.81	0.29	1.86
	Raw	400-2500	0.93	0.47	2.66
		400-700	0.97	0.42	2.47
		701-1000	0.86	0.54	2.58
		1001-2500	1.15	0.22	2.25
Mencía	SNV	400-2500	0.82	0.59	2.89
		400-700	0.92	0.43	2.69
		701-1000*	0.81	0.59	2.52
		1001-2500	0.87	0.54	2.88
	Raw	400-2500	1.51	0.29	2.22
		400-700	1.48	0.23	2.31
		701-1000	1.45	0.26	2.47
		1001-2500	1.59	0.29	2.23
	SNV	400-2500*	1.26	0.47	2.85
		400-700	1.47	0.23	2.41
		701-1000	1.28	0.40	2.86
		1001-2500	1.29	0.45	2.76
Tempranillo	Raw	400-2500	1.51	0.29	2.22
		400-700	1.48	0.23	2.31
		701-1000	1.45	0.26	2.47
		1001-2500	1.59	0.29	2.23
	SNV	400-2500*	1.26	0.47	2.85
		400-700	1.47	0.23	2.41
		701-1000	1.28	0.40	2.86
		1001-2500	1.29	0.45	2.76

Supplementary Material S2

Figure S1 shows the distribution of weighted regression coefficients of the best regression models for the grape varieties. The best results were obtained with SNV-transformed data in the NIR spectral range for Godello, Mencía, and Verdejo, while for Tempranillo was achieved with SNV-transformed data in full spectral range. For white grapes, Godello present minimum and maximum peaks at 750 and 920 nm and 725, 875 and 950 nm, respectively. Verdejo has a minimum peak at 910 nm and maximum at 950 nm. In red grapes, Mencía shows minimum peaks at 840, 915 and 985 nm. The maximum peaks are located at 880, 950 and 1000 nm. Tempranillo in full spectral range, presents diverse minimum peaks at 600, 710 and 1000 nm. The maximum peaks are located at 700, 800, 995 and 1100 nm.

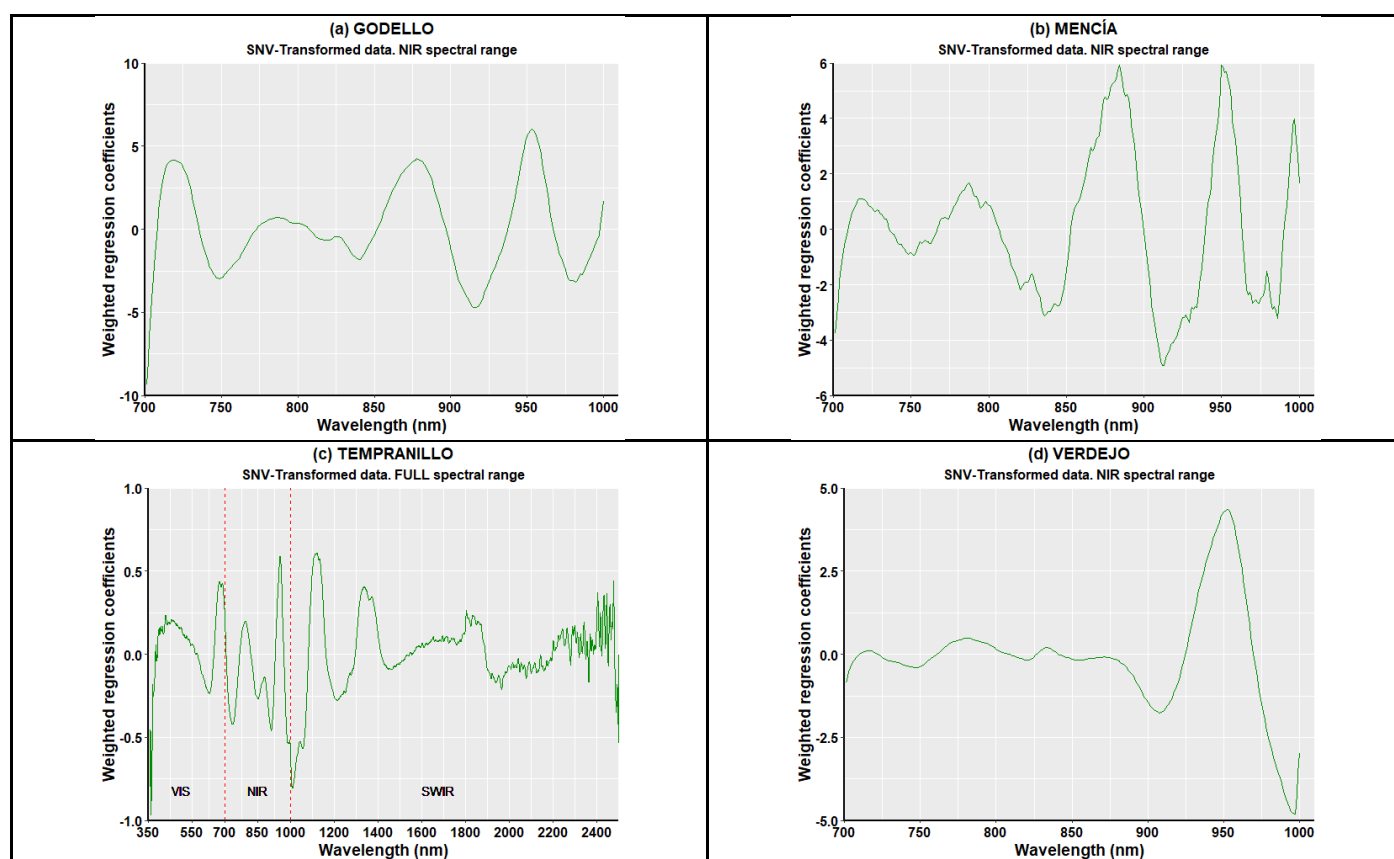


Figure S1. Distribution of weighted regression coefficients of the best regression models for the Godello (a), Mencía (b), Tempranillo (c), and Verdejo (d) grape varieties.

The previous figures have been developed in R Statistical Software (version 4.2.0; R Core Team 2022) and RStudio (version 2022.07.2+576; RStudio Team, 2022). The main packages used were: ggplot2 (v3.3.6; Wickham, 2016), tidyverse (v1.3.1; Wickham, 2019), ggspectra (v0.3.9; Aphalo, 2015) and others.