International Conference of Agricultural Engineering
AgEng 2014 Zurich 6–10 July

Ref: C0509

Elimination of variability sources in NIR spectra for the determination of fat content in olive fruits

Natalia Hernández-Sánchez, Physical Properties Laboratory-Advanced Technologies In Agri-Food (Lpf-Tagralia), Technical University of Madrid, 28040 Madrid, Spain

Beatriz Somoza-Rodríguez, Departamento De Producción Vegetal: Fitotecnia, Technical University of Madrid, 28040 Madrid, Spain

Jean Michel Roger, Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (IRSTEA), Montpellier, France

María Gómez-Del-Campo, Departamento De Producción Vegetal: Fitotecnia, Technical University of Madrid, 28040 Madrid, Spain

Abstract

Models for prediction of oil content as percentage of dried weight in olive fruits were computed through PLS regression on NIR spectra. Spectral preprocessing was carried out by applying multiplicative signal correction (MSC), Savitzky–Golay algorithm, standard normal variate correction (SNV), and detrending (D) to NIR spectra. MSC was the preprocessing technique showing the best performance. Further reduction of variability was performed by applying the Wold method of orthogonal signal correction (OSC). The calibration model achieved a R^2 of 0.93, a SEPc of 1.42, and a RPD of 3.8. The R^2 obtained with the validation set remained 0.93, and the SEPc was 1.41.

Keywords: Near infrared, olive, oil content, OSC

1 Introduction

The determination of fat content in olives fruits is of major interest in the process of olive oil extraction.

The implementation of near infrared (NIR) Spectroscopy in devoted models allows the mills to carry out rapid measurements of oil content with minimum sample preparation. For a prediction model to be robust it is a major requirement that most of the expected sample variability is taken into account when generating the model. Consequently, the variation of the spectral features that is not related to the parameter under study is enhanced, which could require special orthogonalization techniques.

In addition, as a result of the heterogeneity of the solid-like samples the spectra are affected by both multiplicative and additive effects due to scattering. Hence, the application of spectral preprocessing techniques is usually required.

In the present work the major criterion for achieving the maximum variability in the calibration set was based on the different quantity of radiation intercepted by the foliar surface according to the position in the tree and to the hedgerows orientation (Gómez-del-Campo et al., 2009; Connor et al., 2009). In order to enhance the variability others main sources were included

such as season and geographical location of orchards. Different spectral preprocessing techniques were evaluated so as to obtain a highly accurate and robust model.

2 Materials and methods

Olive fruits cv. Arbequina were manually picked from different positions in the tree, hedgerows orientations, seasons and geographical locations (Figure 1). Samples were randomly split up into two sets: calibration (70%) and validation (30%). The oil content was measured from dried samples using an NMR Minispec NMS100 (Bruker Optik GmbH) since the suitability of this analytical technique, in terms of accuracy and precision, has been proved through comparison to the official method based on Soxhlet extraction. Values were expressed as percentage of dried weight and ranged from 32.28% to 54.76%.

Figure 1: Left, layers delimitation in olive tree for sample picking. Right, variability in color of olive skin according to different maturity stages.

Figure 2 highlights the dispersion of the oil content values for both calibration and validation sets; and the representativeness of the validation set according to the dispersion of such values. Table 1 summarizes statistics for both sets.

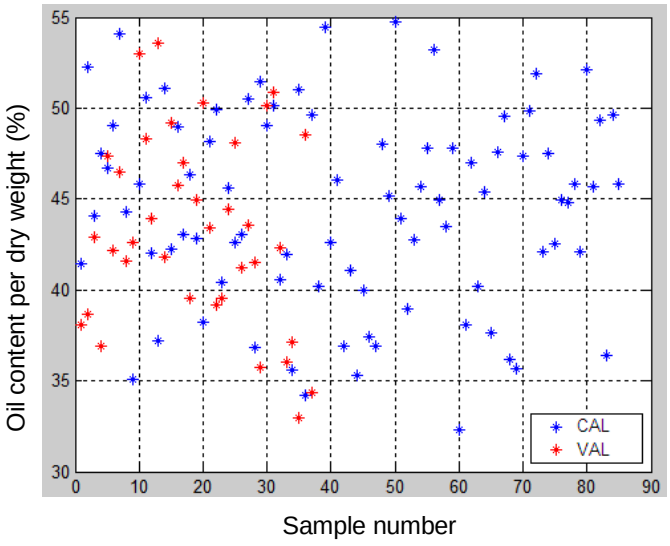

Figure 2: Oil content per dry weight (%) for calibration and validation samples.

Table 1: Calibration and validation sets statistics.

STATISTICS	Calibration set N = 85	Validation set N = 37
Minimum	32.28	32.98
Maximum	54.76	53.55
Range	22.48	20.56
Mean	44.45	43.33
Median	44.91	42.90
STD	5.38	5.26

NIR measurements were performed by a FOSS NIRSystems 5000 in the 1100-2500 nm range at 2 nm resolution in reflectance mode, then transformed to absorbance values as log(1/R).

The identification of the outliers was carried out by univariate and multivariate approaches, for chemical and spectral data respectively.

The chemometric analysis was performed using Matlab 7.0 (The Mathworks Inc.) by applying partial least squares regression to generate the calibration model, which was internally validated through cross-validation. The bias corrected standard error of prediction (SEPc), the determination coefficient (R^2) and the ratio of the standard deviation of the response variable to the SEPc (RPD) were used as parameters of model accuracy.

Spectral preprocessing was carried out by applying the most common techniques (Rinnan et al., 2009) such as multiplicative signal correction (MSC), Savitzky–Golay algorithm, standard normal variate correction (SNV), and detrending (D). The preprocessing technique showing the best performance was used for further reduction of non informative variability.

Thus, remaining variability that was not related to fat content was minimized by applying a dimensionality-reduction method, that is, the Wold method of orthogonal signal correction (OSC) (Wold et al, 1998; Zeaiter et al., 2005). This method computes PCA on the matrix of spectra, then, the matrix of scores is orthogonalized to oil content vector. Latent structures that allow correcting original spectra are obtained by partial least square regression.

3 Results and Discussion

Models performances are summarized in Table 2. Similar performance was found for common techniques, except for SNV. In order to simplify preprocessing computation, MSC was chosen for subsequent OSC prior to PLS regression.

Table 2: Summary of performance results of the PLS models obtained after different preprocessing techniques.

Preprocessing techniques	CALIBRATION				VALIDATION	
	Latent variables	R^2	SEPc	RPD	r^2	SEPc
MSC	**3**	0.80	2.43	2.22	0.83	2.15
SavGol	**3**	0.80	2.42	2.22	0.84	2.13
SNV	**3**	0.78	2.54	2.12	0.86	1.95
SNV+D	**3**	0.80	2.42	2.22	0.84	2.13
SavGol+SNV	**3**	0.81	2.36	2.28	0.84	2.11
MSC + OSC	**3**	**0.93**	**1.42**	**3.79**	**0.93**	**1.41**

The model, comprising 3 latent variables as selected according to comparison between calibration and cross-validation errors of prediction, achieved a R^2 of 0.93, a SEPc of 1.42, and a RPD of 3.8, which means that the prediction is accurate and reliable. The R^2 obtained with the validation set remained 0.93, which is an indication of robustness as well.

Figure 3 shows the increasing reduction of variability between spectra after applying consecutively MSC and OSC corrections (Figure 3, from a to c). When comparing PLS coefficients of the corresponding models (Figure 2, d and e) for oil content prediction, it is shown that spectral regions of interest are reduced, with an increasing importance of peaks around 2312 and 2350 nm, and a decreasing importance of peaks around 1726 and 1762 nm, all of them corresponding to $–CH_2$ bands of oil (Sun, 2009).

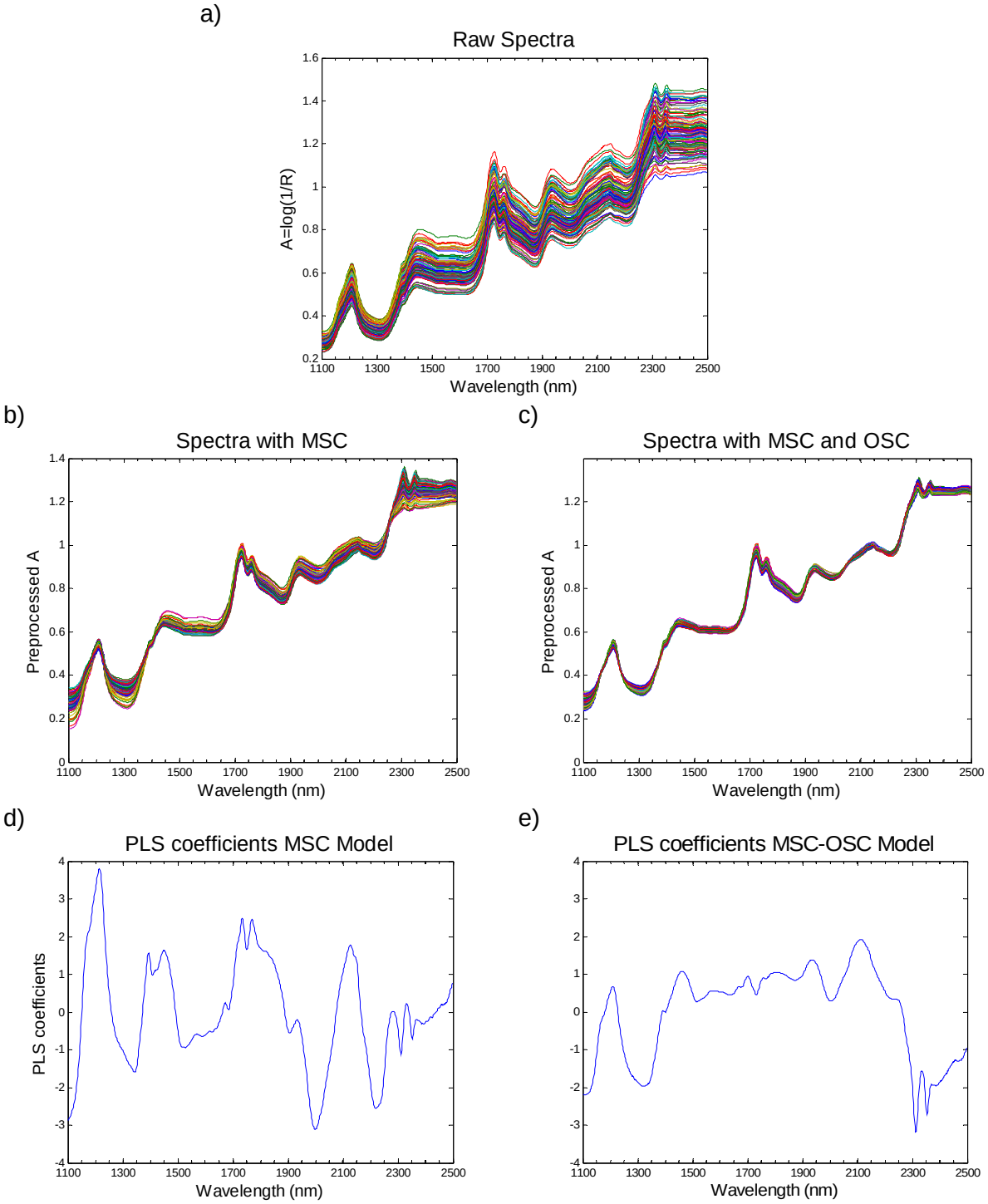

Figure 3: NIR spectra, a) raw NIR spectra; b) NIR spectra after MSC preprocessing; c) NIR spectra after MSC and subsequent OSC preprocessing; d) PLS coefficients of model related to MSC preprocessing; e) PLS coefficients of model related to MSC and subsequent OSC preprocessing.

Further work is foreseen with larger set of samples for both, the generation and validation of reliable and robust models.

Previous published works on olives fruits, fresh and dehydrated, reported lower R^2, higher SEP, and usually higher model complexity as higher number of latent variables were included in the models (Cayuela et al., 2009; Cayuela and Camino, 2010; Salguero-Chaparro et al,

2013; León et al., 2003). The application of OSC preprocessing could improve respective models performances.

4 Conclusions

Orthogonal signal correction is a powerful spectral preprocessing technique that allows the reduction of sources of variability that are not related to the parameter to be predicted, which noticeably enhances the performance of common techniques applied prior to PLS modeling.

5 Acknowledgements

We gratefully acknowledge Jacinto Cabetas from El Carpio de Tajo, Antonio Capitán from Écija, Casas de Hualdo from Puebla de Montalbán and Todolivo from Pedro Abad for access to olive orchards where this research was conducted.

6 References

Cayuela, J. A. and Camino, M. d. C. P. (2010), Prediction of quality of intact olives by near infrared spectroscopy. *Eur. J. Lipid Sci. Technol., 112*, 1209-1217. doi: 10.1002/ejlt.201000372.

Cayuela, J.A., García, J.M., Caliani, N. (2009). NIR prediction of fruit moisture, free acidity and oil content in intact olives. *GRASAS Y ACEITES, 60 (2)*, 194-202.

Gómez-del-Campo, M., Centeno, A., Connor, D.J. (2009). Yield determination in olive hedgerow orchards. I. Yield and profiles of yield components in north–south and east–west oriented hedgerows. *Crop & Pasture Science, 60*, 434-442.

Connor, D.J., Centeno, A., Gómez-del-Campo, M. (2009). Yield determination in olive hedgerow orchards. II. Analysis of radiation and fruiting profiles. *Crop & Pasture Science, 60*, 443-452.

León, L., Rallo, L., Garrido, A. (2003). Análisis de aceituna intacta mediante espectroscopia en el infrarrojo cercano (NIRS): una herramienta de utilidad en programas de mejora de olivo. *Grasas y Aceites, 54 (1),* 41-47.

Rinnan, A., Berg, F., Engelsen, S.B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry, 28 (10).*

Salguero-Chaparro, L., Baeten, V., Fernández-Pierna, J.A., Peña-Rodríguez, F. (2013). Near infrared spectroscopy (NIRS) for on-line determination of quality parameters in intact olives, *Food Chemistry*, doi: http://dx.doi.org/10.1016/j.foodchem.2013.01.002.

Sun, D-W. (2009). Infrared Spectroscopy, for food quality analysis and control. (1st ed.). UK: Academic Press, (Chapter 1).

Svante Wold, S., Antti, H., Lindgren, F., Ohman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems 44*, 175-185.

Zeaiter, M., Roger, J.-M., Bellon-Maurel, V. (2005). Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. *Trends in Analytical Chemistry, 24 (5)*.