Data standardization strategy to correct the effect of seasonality in the authentication of virgin olive oil

Glòria Rovira, Itziar Ruisánchez* and M. Pilar Callao
Chemometrics and Sensorics for Analytical Solutions (CHEMOSENS) Group, Department of Analytical and Organic Chemistry, Rovira i Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain

Abstract

This study proposes a standardization strategy for dealing with seasonal variability in the authentication of extra virgin oils from the PDOs Les Garrigues and Siurana. Samples from four harvests were measured by fluorescence spectroscopy. A PLS-DA two-class model was developed and validated from samples from one of the harvests. When samples from three other harvests were predicted with the model developed, it was observed that the sensitivity and specificity were lower than when the model was validated. For the standardization process, we adapted the PDS technique to obtain the transfer function. The results obtained from the transformed spectra show that standardization is a good strategy for extending the usefulness of the models if the samples to be predicted are subject to seasonal variability.

Keywords

Multivariate standardization; PLS–DA; olive oil authentication, Multivariate screening; Food authentication; Extra-virgin olive oil

1. Introduction

Olive oil is of great interest because of its nutritional value and major health benefits. This has generated considerable research interest in this field. A recent bibliometric review describes the global situation and evolution of olive oil research, considering various aspects (trends in the number of publications and distribution of them by countries and institutions, journals and research areas, analysis of authors, citations and co-citations and analysis of most relevant keywords) [1].

Compared to other types of oil, olive oil is more expensive so it is susceptible to fraud. Recently, this issue has been the subject of an interesting paper [2] that considers various types of fraud, of which two are particularly important: adulteration with other cheaper oils and lack of authenticity due to fraudulent labeling [3].

Since authentication requires a qualitative response (yes or no), the most appropriate methodologies are those that use instrumentation to obtain the multivariate signals and multivariate classification techniques to treat them. A recent review summarizes the use of various analytical techniques coupled with multivariate data analysis to trace the geographical origin of edible oils [4].

In general terms, the main problem involved in establishing and validating a multivariate classification model is finding a large number of representative samples whose class membership is unambiguously known (training/test sets). From them, the main performance parameters, sensibility and specificity are estimated [5], which give information about the model's ability to predict future samples. If the future samples have the same variability as those used to build the model, the model's performance parameters are maintained. However, when the future samples contain new sources of variability, the model may not be suitable for predicting them since they do not fit it, so the model's performance parameters are no longer kept. This might be the case of authenticating seasonal products (or their derivatives) because the samples may contain sources of variability not considered when the model was developed. Therefore, making the model robust to seasonal variability, which involves maintaining the performance parameters, is still an important challenge in the field of authentication nowadays.

To deal with new sources of variability not considered in the training data set, three strategies can be followed:

1) Develop a new model for each season [6]. The main limitation of this strategy is that it cannot take advantage of historical data. In addition, the collection of sufficient training samples for each season can be time-consuming and challenging.

2) Update the model by adding samples of the new conditions, so the variability due to the season is considered. In principle, this alternative seems the most promising, since enough representative samples of the possible variations would eventually be obtained, although it may not be very effective in the first years. Additionally, increasing the sources of variability usually decreases the model's performance parameters.

3) Correct the spectra obtained in the new conditions so that they resemble the spectra of the training set used to build the multivariate model by calculating and applying a transform function. As a result, the model can be used to predict samples measured in the new conditions. It has the advantage of taking historical data into account, the multivariate model need not be built or updated, and few samples are required, the authenticity of which is unambiguously known, for standardization. This approach is implemented by chemometric tools known as standardization or transfer techniques.

The standardization strategy was initially designed to extend the applicability of multivariate calibration NIR models developed with a master instrument to be used by secondary instruments (hence, this type of technique is also known by the alternative name of transfer methods). The first standardization methods introduced by Wang [7] were direct standardization (DS) and piecewise direct standardization (PDS). The latter is still the most widely used. It determines a transformation matrix, F, which relates the spectra in the two conditions. Once established, the spectrum of a sample measured under the second conditions is transformed by applying matrix F so that it resembles the spectra obtained under the first conditions. This transformation allows it to be properly predicted by the model developed in the first conditions.

Most applications of transfer techniques have been developed for IR signals, although they have also been used with other types of multivariate signals, such as UV-visible data [8-10], sensor arrays [11-13], polarography [14], mass spectrometry [15], fluorescence [16-19] and HPLC [20]. They have been used above all in multivariate calibrations [10,21-23], but also to a lesser extent in multivariate classification problems and 2nd order calibrations [16,19+.

Likewise, they have been used to correct for effects other than the change of instrument, for example, the effect of time [10] or temperature (Chen, Morris & Martin, 2005)[24]. Generally, to establish the transformation function F, the same sample has to be measured under both

conditions but this is not always the case and a recent review has focused on studies that determine the transfer function without the need to use the same samples in the two situations considered [25].

The goal of this paper is to propose a standardization strategy for dealing with seasonal variability in the authentication of extra virgin oils from two denominations of origin (Les Garrigues and Siurana). We worked with four data sets from four harvests that were measured using fluorescence spectroscopy. Therefore, in some cases the environmental conditions in each harvest were different. One of these data sets was used to develop a Partial Least Squares Discriminant Analysis (PLS-DA) two-class model for the first conditions.

Since it is impossible for samples from different harvests to be the same, we propose to obtain the transfer function from the average of the spectra of both conditions. This strategy is described in Fig. 1, which schematically shows the steps involved.

Our aim is to show the potential of standardization techniques in classification from olive oil spectroscopic data problems. The strategy is simple and easy to implement in a routine laboratory and it has practical advantages such as maintaining the quality performance of the multivariate model developed.

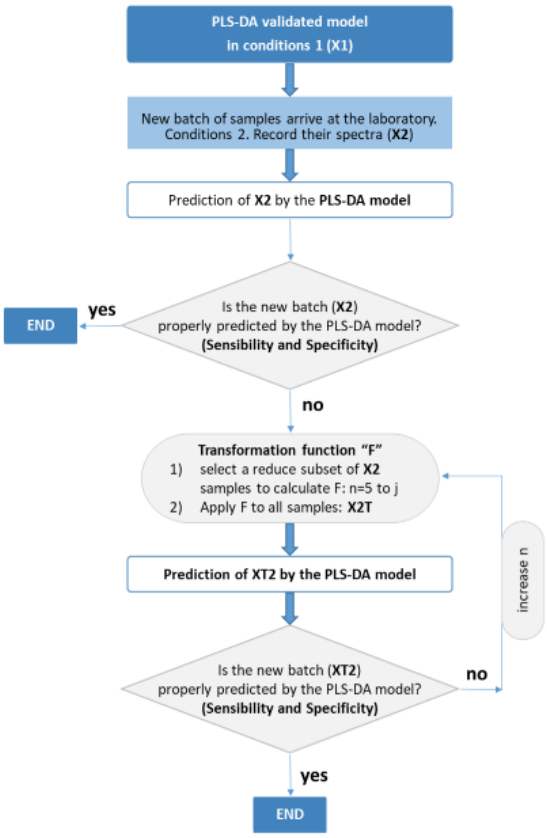


Figure 1. Flow chart of the standardization strategy proposed in this paper.

2. Theoretical background

PLS-DA is a PLS regression technique adapted to a classification technique. It requires two matrices, one with independent variables (matrix **X**), which in our case are the fluorescence spectra, and the other with dependent variables (matrix **Y**), which in our case is a binary code (0 and 1) where 1 indicates sample membership and 0 does not. There is an extensive bibliography on the theoretical and practical aspects of PLS-DA. We have included references to just two recent reviews here, which is by no means exhaustive, but these two references provide many more [26,27].

To validate the classification model, the main performance parameters – sensitivity (SEN) and specificity (SPC) – [5,28] are calculated from the probabilities of the four well-known binary responses [29]: true positive (TP) and true Negative (TN), when the qualitative method rightly considers a sample to be positive or negative; and false positive (FP) and false negative (FN) when the qualitative method wrongly considers a sample to be positive or negative.

The theoretical foundations of various standardization processes can be found in the references [7,8,25]. To carry out the standardization, we propose adapting the PDS technique.
The PDS technique considers that when the experimental conditions change, the spectra may drift in any direction, both vertically (changes in sensitivity) or horizontally, which means a slight shift in the real wavelengths at which the sample absorb.

To apply PDS, a subset of samples measured in first and second conditions is available. From these two submatrices, a multivariate regression (PCR or PLS) is established between the absorbance in first conditions at a wavelength i and a vector of absorbances around wavelength i, in second conditions.

$$\mathbf{a_{1i}} = \mathbf{x_i}^T \mathbf{f_i} + \mathbf{f_{0i}} \quad (1)$$

where $x_i^T=[a_{2(i-j)}, \dots , a_{2(i)}, \dots , a_{2(i+j)}]$, and the length of the vector (2j+1) is known as window size.

The process is repeated for all the wavelengths and n vectors with f coefficients, and n values of f_0 are obtained, which are grouped in the corresponding matrix F

$$F = \mathrm{diag}(\mathbf{f}) \quad (2)$$

In this study, the **f** is established by a linear relation between **v1** and **v2**, as shown in Equation 3:

$$\mathbf{v_1} = \mathbf{v_2} \cdot \mathbf{f} \quad (3)$$

being, **v1** the average of the **X1** matrix containing the spectra of the samples with which the model was developed (harvest in conditions 1). Similarly, **v2** is the average of the **X2** matrix containing the spectra of a subseries of samples from the harvests to be predicted (harvest in conditions 2).

where **f** is the transformation or transfer vector which contains as many values as the spectra dimension (*f1, f2, f3,….fn*), with *n* being the number of wavelengths considered. From **f**, the transformation matrix F is obtained (Equation 2).

Equation 4 is applied so that the spectra of the new conditions (harvest in conditions 2) resemble the spectra of the first conditions:

$$\mathbf{X_{T2} = X_2 \cdot F} \quad (4)$$

where $\mathbf{X_{T2}}$ is the transform matrix of spectra in conditions 2 as if they had been obtained in conditions 1. Therefore, $\mathbf{X_{T2}}$ will be used in the prediction step.

3. Samples, instrumentation, and software

The data set consisted of 330 samples from four different harvests. Table 1 shows a summary of the number of samples from each harvest and each category, Les Garrigues (LG) and Siurana (S). All samples were supplied by the Catalan Government's Official Tasting Panel of Virgin Olive oils of Catalonia, which confirms the authentication of the oils.

Table 1. Number of samples available in each harvest (conditions) and in the two DOs studied.

Conditions	Nº of samples	Les Garrigues	Siurana
1	156	96	60
2A	72	36	36
2B	58	19	39
2C	44	19	25
Total	330	170	160

The two-class PLS-DA model was built with samples from the harvest in conditions 1, the **X1** matrix. The 156 samples available were divided into two subsets: the training and test set. The samples were assigned to the two subsets using the Kennard-Stone algorithm [30]. The criteria determining the number of samples from each class in the training set (50 for Les Garrigues and 44 for Siurana) were that it should be similar for both classes and be around 75% of the samples available.

Fluorescence analysis was carried out using a Shimadzu RF-5301PC (Shimadzu Corporation, Kyoto, Japan). The emission spectra were collected between 360 and 800 nm using an excitation wavelength of 350 nm and a slit width of 5 nm. The integration time was 0.1 nm and the wavelength was increased every 10 nm during the scanning of the spectra.

The spectra obtained were processed and the models were established using MATLAB software, version 8.0.0.783 – R2012b (Natick, MA, USA) and PLS Toolbox 7.0.2 (Eigenvector Research Inc., Wenatchee, WA, USA).

4. Results and discussion

Fig. 2 shows the average spectrum of the samples for the four harvests, and for Les Garrigues (figure 2a) and for Siurana (figure 2b). The major differences with the Siurana spectrum recorded in conditions 1 (black line) were observed in the spectra recorded in conditions 2B and 2C. And for the Les Garrigues spectrum, the differences were observed, above all, in conditions 2A and 2C. Although the mean spectra in the different conditions are expected to present a certain variability, the graphs suggest that the harvest year is a factor that can present new sources of variability.

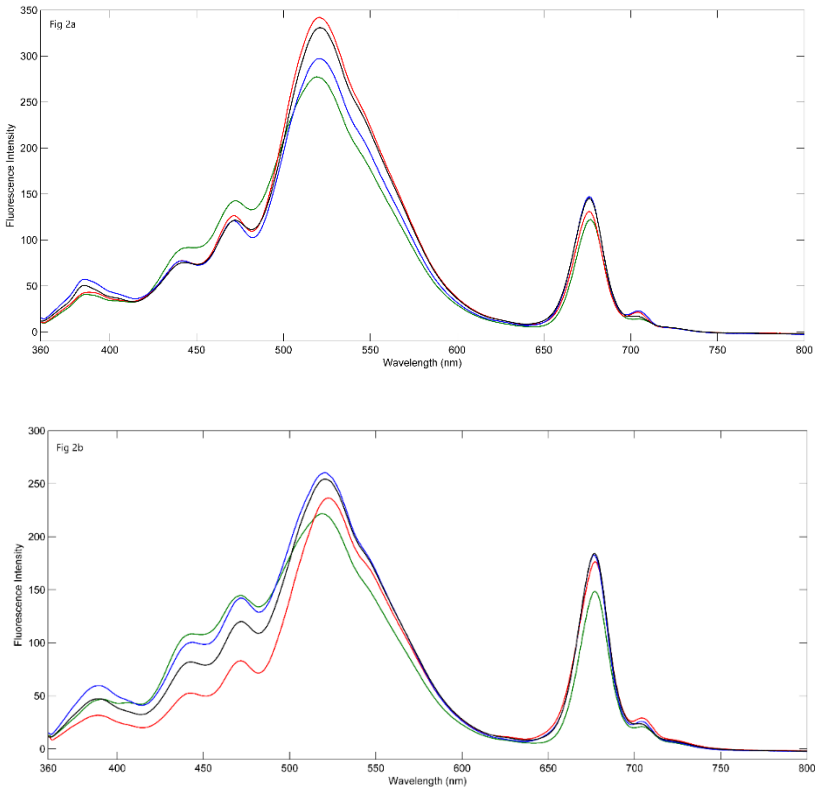


Figure 2. Average spectrum of the samples for the four harvests studied for a) Les Garrigues and b) Siurana. Color code: black for condition 1; blue, red and green for conditions 2A, 2B and 2C, respectively.

The proposed standardization strategy is summarized in figure 1. The first step is to establish and validate the PLS-DA model with samples from the harvest in conditions 1. Before modeling, all samples were pre-treated with baseline correction and centering. The two-class PLS-DA classification model was built for the two DOs (Les Garrigues and Siurana). The model required 5 LVs with an explained variance of around 99%. The second step is to predict the samples from other harvests (in our case, three conditions: 2A, 2B, and 2C) to check the model's ability.

Table 2a shows the confusion matrix. From it, the different performance parameters (Accuracy, False and True Positive rate, False and True Negative rate, Precision, and F-Score) are obtained [5]. Table 2b shows the main performance parameters for each harvest and one class (Les Garrigues). Since it is a discriminant two-class model, the values for class 1 (Les Garrigues) are the same as for class 2 (Siurana) but with the sensitivity and specificity values inverted, that is to say, the sensitivity of the Siurana class is the same as the specificity of the Les Garrigues class. The opposite occurs with the results of the prediction of the Les Garrigues samples.

Table 2a. Confusion Matrix for Les Garrigues (LG) and Siurana (S)

Table 2b. Quality parameters obtained for the Les Garrigues class when the PLS-DA model built with the original data (conditions 1) was used to predict all data sets (conditions 1 and 2).

Table 2a

Condition	TP	FP	FN	TN
1 training (LG)	45		5	
1 training (S)		5		39
1 test (LG)	37		6	
1 test (S)		0		14
2A (LG)	27		9	
2A (S)		0		36
2B (LG)	17		2	
2B (S)		13		26
2C (LG)	18		1	
2C (S)		6		19

Table 2b

Conditions	Sensitivity (%)	Specificity (%)
1 training	90	89
1 test	86	100
2A	**75**	100
2B	90	**67**
2C	95	**76**

It can be seen that the sensitivity of the samples in conditions 2B and 2C is of the order of that of the samples in conditions 1, in both the training and the test set. But the sensitivity in conditions 2A is worse (value in bold). In other words, the samples from the Les Garrigues vintages 2B and 2C are properly recognized by the model as authentic Les Garrigues just as the Les Garrigues samples are recognized by the model in conditions 1. But this is not the case in conditions 2A. The specificity value is notably worse than the values obtained in conditions 1 for

the samples from harvests 2B and 2C (values in bold). Thus, the ability to recognize that the Siurana samples are not from Les Garrigues in these conditions was lower than in conditions 1.

On the basis of these results and following the diagram in figure 1, the samples that have to be transformed are vintages 2B and 2C from Siurana and vintage 2A from Les Garrigues since these new batches (X2) were not properly predicted by the PLS-DA model.

The next step (Fig. 1) is to establish the transformation function "F" (equation 2). The transfer vector "**f**" is calculated by selecting a smaller subset of **X2** samples to find the average spectrum in the second conditions. To start with, a small number of samples should be selected and then, if necessary, gradually increased. For greater representativeness in the results, the selection has been carried out in triplicate, so the samples have been selected randomly and not following a pre-established criterion such as Kennard-Stone, for example. So, as a result, three independent vectors were obtained for each class and for each one of the second conditions that are to be corrected ($\mathbf{v2A_{LG,3s}}$; $\mathbf{v2B_{S,3s}}$; and $\mathbf{v2C_{S,3s}}$). By increasing the number of selected samples, new vectors were obtained under conditions 2 such as $\mathbf{v2A_{LG,\,nS}}$, etc.

Similarly, the vector for the average values of the spectra of harvest conditions 1 were obtained for each category ($\mathbf{v1_{LG}}$, and $\mathbf{v1_{S}}$). The transfer vector "**f**" was obtained by relating **v1** with the corresponding vectors under conditions 2, for instance $\mathbf{v1_{LG}}$ with $\mathbf{v2A_{LG,3S}}$ (equation 1). Then, the corresponding transfer matrices "**F**" were obtained by applying equation 2. Finally, equation 3 was applied to obtain the transformed data matrices $\mathbf{XT_{2A,LG}}$, $\mathbf{XT_{2B,S}}$ and $\mathbf{XT_{2C,S}}$ in triplicate, one for each of the three random replicates used to select the number of samples for the standardization.

Table 3a shows the confusion matrix obtained from the standardized spectra. Table 3b shows the performance parameters obtained when predicting the transformed matrices with the PLS-DA model built in conditions 1. We will regard the results as satisfactory if the values of these parameters are similar to those obtained under conditions 1. A comparison of the results of transforming the data from the average vector with three samples with the results of no transformation (Table 2) shows that the sensitivity of transformed condition 2A is higher than that obtained with the untransformed data, but not as high as that obtained under conditions 1 in two of the three random standardizations. Following the procedure proposed in the scheme of figure 1, the same process is carried out but with five samples. The results are shown below in the same table and the values are satisfactory.

Table 3a. Confusion Matrix after standardization for Les Garrigues (LG) ans Siurana (S)

Conditions	Nº of samples	Random selection	TP	FP	FN	TN
2A	3	1	28		5	
		2	33		0	
		3	28		5	
	5	1	26		5	
		2	30		1	
		3	25		6	
2B	3	1		2		34
		2		2		34
		3		2		34
	5	1		1		33
		2		1		33
		3		2		32
2C	3	1		0		22
		2		1		21
		3		7		15
	5	1		0		20
		2		3		17
		3		1		18

Table 3b. Prediction of the results of the transformed spectra in three different conditions (2A, 2B, and 2C) with the model established with spectra in conditions 1.

Conditions	Nº of samples	Random selection	Sensitivity (%)	Specificity (%)
2A	3	1	81	
		2	100	
		3	81	
	5	1	90	
		2	94	
		3	97	
2B	3	1		94
		2		94
		3		94
	5	1		97
		2		97
		3		94
2C	3	1		100
		2		91
		3		97
	5	1		100
		2		85
		3		90

For the Siurana conditions 2B and 2C, with the standardization starting from three samples, results were satisfactory. Although the scheme presented in figure 1 suggests that three samples are sufficient to obtain the average vector, we also tested with five samples to observe the effect of the number of samples. It can be seen that the more samples there are, the greater the quality of the parameters obtained. But it should be pointed out that this improvement was not significant, since the results cannot be expected to be better than the results obtained in the original conditions in which the model was established (conditions 1).

Figure 3 shows the mean spectra under conditions 1, under conditions 2, and under conditions 2 after standardization (**XT2**) with different numbers of randomly selected samples. Figure 3a shows the spectra from Les Garrigues under conditions 1 and 2A, figure 3b the spectra from Siurana under conditions 1 and 2B, and figure 3b the spectra from Siurana under conditions 1 and 2C. In all cases, as might be expected, the transformed spectra under conditions 2 were similar to, but not exactly the same as, the spectra under conditions 1. We should also point out that there were no patterns in terms of the number of samples used. Therefore, if a small number of standardization samples provides satisfactory results in terms of the quality parameters, the process can be ended.

Figure 3. Average spectrum of samples measured in original conditions and after the standardization using different numbers of randomly selected samples:

Figure 3a Les Garrigues, conditions 1 (black solid), conditions 2A (blue solid) and conditions 2A after standardization with three samples (blue dashes) and with five samples (blue dots);

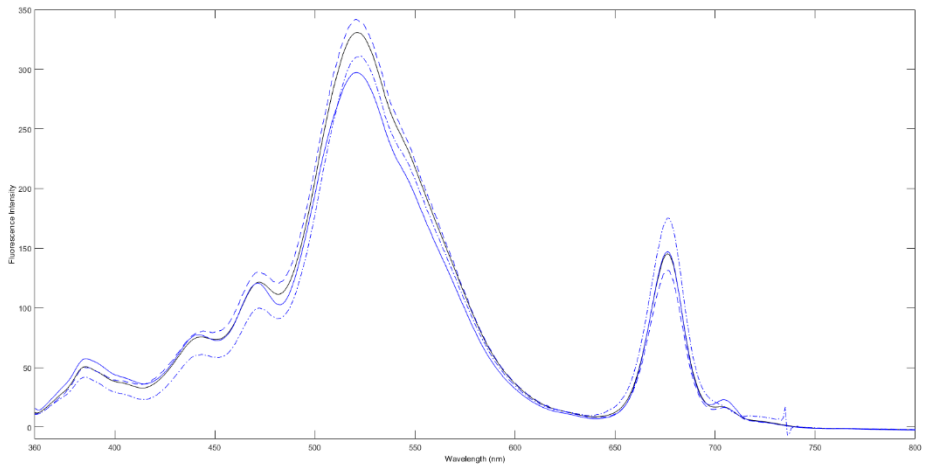


Figure 3b. Siurana, conditions 1 (black solid), conditions 2B (red solid), and conditions 2B after standardization with three samples (red dashes) and with five samples (red dots);

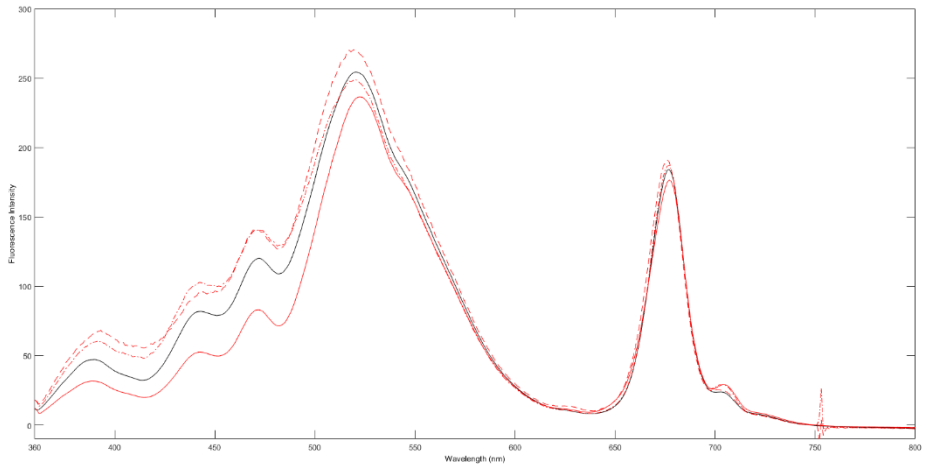
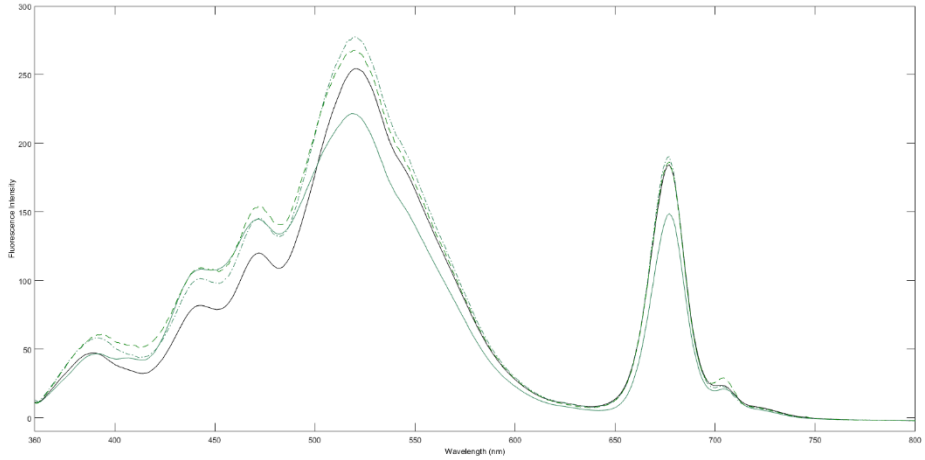


Figure 3c. Siurana, conditions 1 (black solid), conditions 2C (green solid), and conditions 2C after standardization with three samples (green dashes) and with five samples (green dots)



5. Conclusions

A two-class PLS-DA classification model was developed to classify virgin olive oil from two Catalan PDOs (Les Garrigues and Siurana). The model was developed with samples from one harvest using fluorescence measurements.

When samples were predicted from three other harvests, the quality parameters (sensitivity and specificity) were lower since the seasonal variability was not the same as the one used in the model. Therefore, the PDS standardization technique was adapted to extend the usefulness of the model. As a general trend, the results of the standardization process are comparable to those obtained from the initial PLS-DA model.

A great advantage of the proposed standardization strategy is that can be implemented with a small number of samples. From a practical point of view, it is useful when it is difficult to obtain samples that are known to belong to a particular class.

Standardization methods have proved to be valuable tools for preserving the performance of a multivariate classification model when the samples to be predicted are subject to new sources of variability, particularly the effect of seasonality on agricultural products and/or their derivatives.

Acknowledgements

This study was supported by the research program "Program of research activity (2020PMF-PIPF) at the Rovira i Virgili University, Tarragona, Spain. Special mention should be made of M. Angels Calvo, head of the Official Tasting Panel of Virgin Olive Oils of the Catalan Government.

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

[1] Shao, J., Huang, X., Liu, J., Di, D. (2022). Characteristics and trends in global olive oil research: A bibliometric analysis. International Journal of Food Science and Technology, 57, 3311-3325. https://doi.org/10.1111/ijfs.15659.

[2] Yan, J., Erasmus, S. W., Toro, M. A., Huang, H., van Ruth S. M. (2020). Food fraud: Assessing fraud vulnerability in the extra virgin olive oil supply chain. Food Control, 111, 107081. https://doi.org/10.1016/j.foodcont.2019.107081.

[3] Callao, M. P., Ruisánchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. Food Control, 86, 283-293. https://doi.org/10.1016/j.foodcont.2017.11.034.

[4] Tahir, H. E., Arslan, M., Mahunu, G. K., Mariod, A. A., Hashim, S. B. H., Xiaobo, Z., Jiyong, S., El-Seedi, H. R., Musa, T. H. (2022). The use of analytical techniques coupled with chemometrics for tracing the geographical origin of oils: A systematic review (2013-2020). Food Chemistry, 366, 130633. https://doi.org/10.1016/j.foodchem.2021.130633.

[5] López, M. I., Callao, M. P., Ruisánchez, I. (2015). A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. Analytica Chimica Acta, 891, 62-72. https://doi.org/10.1016/j.aca.2015.06.032

[6] Rungpichayapichet, P., Mahayothee, B., Nagle, M., Khuwijitjaru, P., Müller, J. (2016). Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango. Postharvest Biology and Technology, 111, 31-40. https://doi.org/10.1016/j.postharvbio.2015.07.006.

[7] Wang, Y., Veltkamp, D. J., Kowalski, B. R. (1991). Multivariate instrument standardization. Analytical Chemistry, 63, 2750−2756. https://doi.org/10.1021/ac00023a016.

[8] Sales, F., Callao, M. P., Rius, F. X. (1997). Multivariate standardization techniques using UV-Vis data. Chemometrics and Intelligent Laboratory Systems, 38, 63-73. https://doi.org/10.1016/S0169-7439(97)00051-8.

[9] Sales, F., Callao, M. P., Rius, F. X. (2000). Standardization of a multivariate calibration model applied to the determination of chromium in tanning sewage. Talanta, 52, 329-336. https://doi.org/10.1016/S0039-9140(00)00366-0.

[10] Di Anibal, C. V., Ruisánchez, I., Fernández, M., Forteza, R., Cerdà, V., Callao, M. P. (2012). Standardization of UV–visible data in a food adulteration classification problem. Food Chemistry, 134, 2326–2331. https://doi.org/10.1016/j.foodchem.2012.03.100

[11]Sales, F., Callao, M. P., Rius, F. X. (1999). Multivariate standardization techniques on ion-selective sensor arrays. Analyst, 124, 1045-1051. https://doi.org/10.1039/A902585E.

[12] Galvan, D., Bona, E., Borsato, D., Danieli, E., Killner, M. H. M. (2020). Calibration transfer of partial least squares regression models between desktop nuclear magnetic resonance spectrometers. Analytical Chemistry, 92, 12809-12816. https://doi.org/10.1021/acs.analchem.0c00902.

[13] Lindner, S., Burguer, R., Rutledge, D. N., Do, X. T., Rumpf, J., Diehl, B. W. K., Schulze, M., Monakhova, Y. B. (2022). Is the calibration transfer of multivariate calibration models between high- and low-field NMR instruments possible? A case study of lignin molecular weight. Analytical Chemistry, 94, 3997-4004. https://doi.org/10.1021/acs.analchem.1c05125.

[14] Herrero, A., Ortiz, M. C. (1997). Multivariate calibration transfer applied to the routine polarographic determination of copper, lead, cadmium and zinc. Analytica Chimica Acta, 348, 51-59. https://doi.org/10.1016/S0003-2670(97)00154-2.

[15] Coleman, M. D., Brewer, P. J., Smith, J. M., Harris, P. M., Clift, M. G., Milton, M. J. T. (2007). Calibration transfer strategy to compensate for instrumental drift in portable quadrupole mass spectrometers. Analytica Chimica Acta, 601, 189-195. https://doi.org/10.1016/j.aca.2007.08.031.

[16] Sun, X-D., Wu, H-L., Chen, Y., Chen, J-C., Yu, R-Q. (2019). Chemometrics-assisted calibration transfer strategy for determination of three agrochemicals in environmental samples: Solving signal variation and maintaining second-order advantage. Chemometrics and Intelligent Laboratory Systems, 194, 103869. https://doi.org/10.1016/j.chemolab.2019.103869.

[17] Alves, J. C. L., Poppi, R.J. (2013). Pharmaceutical analysis in solids using front face fluorescence spectroscopy and multivariate calibration with matrix correction by piecewise direct standardization. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 103, 311-318. https://doi.org/10.1016/j.saa.2012.10.074.

[18] Nørgaard, L. (1995). Direct standardisation in multi wavelenght fluorescence spectroscopy. Chemometrics and Intelligent Laboratory Systems, 29, 283-293. https://doi.org/10.1016/0169-7439(95)80103-G.

[19] Milanez, K. D. T. M., Nóbrega, T. C. A., Nascimento, D. S., Insausti, M., Pontes, M. J. C. (2017). Transfer of multivariate classification models applied to digital images and fluorescence spectroscopy data. Microchemical Journal, 133, 669-675. https://doi.org/10.1016/j.microc.2017.03.004.

[20] Chen, Y., Wu, H-L., Wang, T., Liu, B-B., Ding, Y-J., Yu, R-Q. (2022). Piecewise direct standardization assisted with second-order calibration methods to solve signal instability in high-performance liquid chromatography-diode array detection systems. Journal of Chromatography A, 1667, 462851. https://doi.org/10.1016/j.chroma.2022.462851.

[21] Myles, A. J., Zimmerman, T. A., Brown, S. D. (2006). Transfer of multivariate classification models between laboratory and process near-infrared spectrometers for the discrimination of green Arabica and Robusta coffee beans. Applied Spectroscopy, 60, 1198–1203. https://doi.org/10.1366/000370206778664581.

[22] Silva, N. C., Pimentel, M. F., Honorato, R. S., Talhavini, M., Maldaner, A. O., Honorato, F. A. (2015). Classification of Brazilian and foreign gasolines adulterated with alcohol using infrared spectroscopy. Forensic Science International, 253, 33–42. https://doi.org/10.1016/j.forsciint.2015.05.011.

[23] Milanez, K. D. T. M., Silva, A. C., Paz, J. E. M., Medeiros, E. P., Pontes, M. J. C. (2016). Standardization of NIR data to identify adulteration in ethanol fuel. Microchemical Journal, 124, 121-126. https://doi.org/10.1016/j.microc.2015.08.013.

[24] Chen, Z-P., Morris, J., Martin, E. (2005). Correction of temperature-induced spectral variations by loading space standardization. Analytical Chemistry, 77, 1376-1384. https://doi.org/10.1021/ac040119g.

[25] Mishra, P., Nikzad-Langerodi, R., Marini, F., Roger, J. M., Biancolillo, A., Rutledge, D. N., Lohumi, S. (2021). Are standard sample measurements still needed to transfer multivariate calibration models between near-infrared spetrometers? The answer is not always. TrAC Trends in Analytical Chemistry, 143, 116331. https://doi.org/10.1016/j.trac.2021.116331.

[26] Biancolillo, A., Marini, F., Ruckebusch, C., Vitale, R. (2020). Chemometric strategies for spectroscopy-based food authentication. Applied Sciences, 10, 6544. https:// doi.org/10.3390/app10186544.

[27] Oliveri, P., Malegori, C., Mustorgi, E., Casale, M. (2021). Qualitative pattern recognition in chemistry: theoretical background and practical guidelines. Microchemical Journal, 162, 105725. https://doi.org/10.1016/j.microc.2020.105725.

[28] Cuadros-Rodríguez, L., Pérez-Castaño, E., Ruiz-Samblás, C. (2016). Quality performance metrics in multivariate classification methods for qualitative analysis. TrAC Trends in Analytical Chemistry, 80, 612-624. https://doi.org/10.1016/j.trac.2016.04.021

[29] Ellison, S. L. R., Fearn, T. (2005). Characterising the performance of qualitative analytical methods: statistics and terminology. TrAC Trends in Analytical Chemistry, 24, 468-476, https://doi.org/10.1016/j.trac.2005.03.007.

[30] Kennard R.W., Stone L.A., Computer aided design of experiments, Technometrics, 11 (1969) 137–148, https://doi.org/10.1080/00401706.1969.10490666.