



# OPEN Determination of the main phenolic compounds of olive (*Olea europaea* L.) leaves by near infrared spectroscopy (NIR)

Jorge F. Escobar-Talavera, María Esther Martínez-Navarro, Gonzalo L. Alonso & Rosario Sánchez-Gómez✉

The main phenolics compounds in *Olea europaea* L. leaves are oleuropein, hydroxytyrosol, hydroxytyrosol hexoside, verbascoside and apingenin-7-glucoside, being oleuropein the most abundant. In the literature, there are different methods to determine some of these compounds, which imply an analytical preparation for their determination. In this line, the use of techniques based on infrared spectroscopy is presented as a promising alternative. Therefore, a near-infrared spectroscopy (NIR) calibration, validation and prediction to determine the main metabolites in olive leaves has been developed. To achieve this, olive leaves samples were analysed by HPLC–DAD to quantify the concentrations of the main phenolic metabolites. Subsequently, a predictive model was developed through NIR calibration, achieving a good  $R^2$  value in all cases, being higher than 0.9 and close to 1 in all calibrations, as the same for  $R^2_{cv}$  in all validations, with an excellent value for oleuropein ( $R^2_{cv} = 93\%$ ). The RPD values were close to 4 for oleuropein and verbascoside and higher than 4 for hydroxytyrosol hexoside and apingenin-7-glucoside, indicating that the model has good reliability. So, good predictive models have been developed to determine the content of the main phenolic compounds in *Olea europaea* L. leaves, not only for the most abundant, but also for others with important bioactive properties.

**Keywords** Apingenin-7-glucoside, NIR calibration, Hydroxytyrosol, Hydroxytyrosol hexoside, Oleuropein, Olive leaves, Verbascoside

The olive tree (*Olea europaea* L.) has a great extraordinary importance in global agriculture and economy, especially in Mediterranean regions where it is widely cultivated for its fruits, which are used to produce olive oil. This product is not only an essential component of the Mediterranean diet, renowned for its health benefits, but also a highly demanded product in the global market, valued for its nutritional and organoleptic properties<sup>1</sup>. In countries like Spain, Italy, and Greece, the olive crop and the olive oil production constitute a significant part of the agricultural economy, generating substantial income and employment in these regions. In Spain, the world's largest olive oil producer, the olive grove area covers 2.75 million hectares according to the Ministry of Agriculture, Fisheries and Food.

However, the olive industry also generates a considerable amount of waste, among which olive leaves stand out. These residues are by-products of pruning and harvesting, and their management represents a logistical and environmental challenge due to the large quantities produced. It is estimated that for every ton of olives produced, around 1.25 tons of olive leaves waste are generated<sup>2</sup>. Traditionally, these by-products have been managed as agricultural waste, with little valorization.

Recently, there has been growing interest in the valorization of these plant residues due to their content of bioactive compounds, among which phenolic compounds are prominent<sup>3</sup>. Among these, oleuropein is one of the most abundant and it is known for its potent antioxidant and anti-inflammatory properties, significantly contributing to the health benefits associated with olive products<sup>4</sup>. Hydroxytyrosol, a product of oleuropein hydrolysis, is another important compound, recognized for its superior antioxidant capacity and cardioprotective effects<sup>5</sup>. Verbascoside, also present in the leaves, has anti-inflammatory and antimicrobial properties, enhancing its potential for pharmaceutical applications<sup>6</sup>. Additionally, hydroxytyrosol hexoside and apingenin-7-glucoside,

Cátedra de Química Agrícola, E.T.S.I. Agronómica y de Montes y Biotecnología (ETISAMB), Universidad de Castilla-La Mancha, Avda. de España s/n, 02071 Albacete, Spain. ✉email: rosario.sgomez@uclm.es

although present in smaller quantities, offer additional benefits: the former enhances the antioxidant properties of hydroxytyrosol, and the latter is known for its anti-inflammatory and anticancer effects<sup>7</sup>. These well-known properties (antioxidant, anti-inflammatory, and antimicrobial among other) are opening possibilities for their use in various industries, from pharmaceuticals to food<sup>8</sup>.

To efficiently harness these phenolic compounds, it is crucial to have precise and rapid methods for their identification and quantification in olive leaves. In this context, the development of a calibration method using Near-Infrared Spectroscopy (NIR) presents a promising solution since NIR technology could offer a non-destructive and rapid alternative for analyzing phenolic compounds, facilitating their study and potential industrial use. Although the bibliography on this subject is limited, a calibration method for oleuropein in olive leaves has already been described using MIR (Mid Infrared spectroscopy)<sup>9</sup>, but not for the other phenolic compounds (hydroxytyrosol, verbascoside, hydroxytyrosol hexoside and apigenin-7-glucoside). In this line, NIR technique has been used in olive leaves to predict models for mineral content quantification<sup>10</sup> and in olive oil to measure its quality in terms of fatty acid content and other parameters such as total polyphenol content, where oleuropein content was considered<sup>11,12</sup>.

NIR analysis offers several advantages over MIR ones, such as greater penetration depth into samples and lower operational costs, but, above all, and compared to chromatographic techniques, it is a rapid and non-destructive methodology that requires less sample preparation, so this could provide a significant advantage when applying it to industry.

For all the above, the aim of this work was to develop a specific NIR calibration method for the quantification of phenolic compounds in olive leaves (oleuropein, hydroxytyrosol, verbascoside, hydroxytyrosol hexoside and apigenin-7-glucoside), providing a useful tool for the research and valorization of these agricultural residues.

## Materials and methods

### Chemicals and reagents

Solvents and reagents employed were all HPLC purity or analytical grade. Acetonitrile (ACN) (CAS: 75-05-8) was supplied from Panreac (Barcelona, Spain). Ultrahigh-purity water was produced using a Milli-Q system (Millipore, Bedford, MA).

The standards of oleuropein (CAS: 32619-42-4), verbascoside (CAS: 61276-17-3), hydroxytyrosol (CAS: 10597-60-1), and apigenin-7-glucoside (CAS: 578-74-5) were obtained from Sigma-Aldrich (Madrid, Spain).

### Plant material

A total of 266 samples of olive leaves (*Olea europaea* L., var. Cornicabra) were obtained during the 2022 and 2023 collection campaigns from the three Designations of Origin (DO) for Extra Virgin Olive Oil (Montes de Toledo, Campo de Calatrava and Campo de Montiel), and also from the “Marca Colectiva Gráfica de la Asociación Mesa del Aceite Sierra de Alcaraz”, which are located in areas included in the Integrated Territorial Investment (ITI) of Castilla-La Mancha Spanish region (Supplementary Table S1). The leaves of each sample were collected from different trees, considering them as independent samples. The leaves were air-dried at room temperature ( $18 \pm 3$  °C) for 3 days. Once dried, the samples were processed according to the method of Martínez-Navarro et al.<sup>13</sup>. Briefly, dried leaves were ground in a knife mill (ARES FML-2000; Filtra Vibración, Barcelona Spain) and sieved using a 0.5 mm mesh, until at least 95% of the total weight of the material must have passed through this sieve.

A moisture content analysis was conducted on the samples using a moisture balance equipped with a halogen lamp, specifically the XM-120 T model (Cobos, Barcelona, Spain) operating at a temperature of 105 °C. When moisture loss was less than 0.1% in 180 s, it was considered that the samples had reached constant mass. The final moisture data of the samples can be seen in Supplementary Table S2.

### Preparation of extracts of *Olea europaea* L. leaves

The extracts preparation was carried out according to the method proposed by Martínez-Navarro et al.<sup>13</sup>. For this purpose, 50 mg of olive leaf powder were weighed and extracted with 25 mL of distilled water using a domestic microwave oven (MS-2819W; Saivod, Madrid, Spain) at 800W for 30 s.

Subsequently, samples were centrifuged at 3500 rpm for 10 min, and the supernatant was filtered through a nylon syringe filter (25 mm diameter; Membrane Solutions, Plano, EEUU) with a pore size of 0.2 µm. All samples were extracted in triplicate.

### Chromatographic conditions

The chromatographic conditions followed those described by Martínez-Navarro et al.<sup>13</sup>. Subsequently, 20 µL of the filtered samples were injected into an Agilent 1200 Series High-Performance Liquid Chromatograph (Agilent, Palo Alto, CA, USA), equipped with a Diode Array Detector (G1315D; Agilent) and coupled to a ChemStation data-processing station, version B.03.01 (Agilent). Separation was achieved using a reverse-phase C18 column, Brisa LC2 (250 mm × 4.6 mm, 5 µm particle size), obtained from Teknokroma (Barcelona, Spain), and maintained at 30 °C.

The mobile phase consisted of water (solvent A) and acetonitrile (solvent B). The elution gradient for solvent B was programmed as follows: 0 min, 5%; 10 min, 20%; 15 min, 30%; 18 min, 30%; 36 min, 50%; 42 min, 100%; 44 min, 100%, 48 min, 5%; and 49 min, 5%. The flow rate was set at 1 mL min<sup>-1</sup>.

All compounds detection was carried out by means of DAD detector, with monitoring wavelength set at 280 nm, comparing the UV-Vis spectra and retention times of the phenolic compounds with those of their corresponding pure standards (Sigma-Aldrich, Steinheim, Germany). Also, the quantification was done through the external standard method based on the calibration curve of the different standards at five different

concentrations achieved using UV–Vis signal ( $R^2=0.99$ ). For hydroxytyrosol hexoside, quantification has been done in terms of hydroxytyrosol. All samples were analyzed in triplicate: one of each extraction.

### Near-infrared spectroscopy measurements

The powdered samples of *Olea europaea L.* leaves were spread to fully cover the surface of a 1 cm<sup>2</sup> quartz Petri dish (Perkin-Elmer, Norwalk, USA) and examined using a Perkin Elmer Spectrum One FT-NIR device (Norwalk, USA) coupled with a near-infrared reflectance accessory (NIRA). Data were collected over a wavelength range of 13333.33–4000 cm<sup>-1</sup> with a fixed resolution of 16 cm<sup>-1</sup>. Each sample was scanned twice, so a total of 532 spectra were obtained.

### Data analysis

The Spectrum-One software version 1.0 was used to perform the pretreatment and select models. To perform the data processing for the calibration model, a multivariate approach was used. Through Partial Least Squares Regression (PLS-R) modelling, noise and variability in the data are reduced, condensing them into a set of relevant latent variables. To improve the quality of the spectral data, Multiplicative Scatter Correction (MSC) pre-processing was used<sup>14</sup>. Outlier removal was applied attending to the samples with concentrations outside the linear range of the calibration lines performed with HPLC–DAD. Smoothing (5 and 13 points) was applied to the spectra as well as the first derivative of Savitzky-Golay (5 and 9 points) for all measurements to increase the robustness of the model.

Due to the fact that two calibration models were tested, Calibration–Validation–Prediction 1 (CVP1) and Calibration–Validation–Prediction 2 (CVP2), the data were divided into different sets: (a) Set 1+2, which included all of the samples analyzed; a) Set 1, which contains 85% of the total of all samples; (b) Set 2, accounting the remaining 15% of the samples. For Set 1 and Set 2, the selection of the samples corresponding to each percentage was random. Set 1+2 represented calibration model 1 and was referred to as CVP1, whereas Set 1 and Set 2 represented calibration model 2 and was referred to as CVP2. The two different procedures (CVP1 and CVP2) were performed to achieve the calibration of the method. First, Set 1+2, containing the whole set of samples, was used for calibration, validation and prediction in a similar way to other NIR studies<sup>9,15,16</sup>, where all calibration steps were performed using the same set of samples, as in the study of Fernandez et al.<sup>17</sup>. Second, the process was repeated by performing calibration and validation with Set 1 and prediction with Set 2, as in the study of Suchat et al.<sup>19</sup>

Different indices were calculated to assess the performance the reliability of the model.  $R^2_{cv}$ : coefficient of determination for cross-validation. RPD: calculated according to Luo et al.<sup>18</sup>.  $RPD_p$ : calculated according to Suchat et al.<sup>19</sup>, and represents the relationship between the standard deviation of the reference data and the standard error of cross-validation or test set validation, but specifically using the predicted reference values instead of the actual. RMSEC: Root Mean Square Error of Calibration, calculated according to Luo et al.<sup>18</sup> RMSECV: Root Mean Square Error of Cross-Validation, determined in a similar way as the RMSEC one, but using the cross-validation data. RMSEP: Root Mean Square Error of Prediction.

Validation was performed through the cross-validation process, thus determining the number of latent variables to consider in adjusting the equation. The  $RPD_c$  index was determined according to Suchat et al.<sup>19</sup>, which relates the standard deviation of the samples to the standard error obtained in cross-validation, indicating the predictive ability of the model. The cross-validation process was used because is more appropriate when a small dimension data set is used<sup>20</sup>.

## Results and discussion

### *Olea europaea L.* leaves metabolite content (HPLC–DAD) for set of samples

*Olea europaea L.* leaves (266) sampled from the different localities included under the D.O. of olive oil from the Spanish region of Castilla-La Mancha were of two campaigns (2022–2023) and submitted to different agronomic practices (intensive, extensive and ecological) and crop system (irrigation and non-irrigation) to obtain a more representative trial. The descriptive statistic data for the main phenolic metabolites content (oleuropein, hydroxytyrosol, hydroxytyrosol hexoside, verbascoside and apigenin-7-glucoside) are shown in Table 1.

Table 1 presents the descriptive statistics of phenolic compounds in leaves samples used to develop the NIR calibrations, such as the range of concentrations found for each compound, as well as their mean, standard deviation (SD), and coefficient of variation (SD/Mean), in percentage, which provides a measure of relative variability. In general, and as shown in Table 1, a wide-ranging concentration value was found in the olive leaves for each phenolic compounds, indicating a good scattering for such model development. It can be observed that the range of oleuropein concentration obtained by HPLC–DAD for the different olive leaves was quite wide, reaching concentrations of up to 171 g/kg which would represent approximately 17% of the total olive leaf. This value represents a high concentration compared to the concentrations determined by Martínez-Navarro et al.<sup>13</sup> for Arbequia variety, where the highest concentration found was 91.0 g/kg. In the work of Ansari et al.<sup>21</sup>, the maximum oleuropein concentration determined was 13 g/kg, whereas Ortega-García and Peragón<sup>22</sup> detected a 60 g/kg, both values being lower than the mean oleuropein concentration obtained in the present study.

In order of abundance, verbascoside was the second most concentrated compound in the leaf samples analysed (Table 1). When compared to Ansari et al.<sup>21</sup> research, it was observed that the verbascoside content was higher than the previous work, and slightly lower than the highest concentration detected by Martínez-Navarro et al.<sup>23</sup>. Other studies<sup>24</sup> observed that the verbascoside content was in the same concentration range as that found in the present research. Such results could be justified by the differences in metabolite levels resulting from the different origins of the samples.

The concentration range found for hydroxytyrosol hexoside compound was from 0.69 to 5.51 g/kg (Table 1). These levels were higher than those found for the Picual variety and for wild olive leaves, which showed

Phenolic compound	Number of samples	N (%)	Range (g/kg leaf dry weight)	Mean	SD	CV (%)
Oleuropein	Set 1 + 2	100	8.55–171	88.52	43.30	48.90
	Set 1	85	8.55–171	101.10	10.72	10.60
	Set 2	15	62.20–168	118.71	2.14	1.80
Hydroxytyrosol	Set 1 + 2	100	0.09–2.40	0.73	0.54	73.97
	Set 1	85	0.09–2.40	0.91	0.77	84.61
	Set 2	15	0.09–2.11	0.99	0.68	68.68
Hydroxytyrosol hexoside	Set 1 + 2	100	0.69–5.51	1.81	0.79	43.64
	Set 1	85	0.69–5.51	1.82	0.25	13.73
	Set 2	15	0.69–2.88	1.98	0.65	32.82
Verbascoside	Set 1 + 2	100	2.54–24.60	6.78	3.76	55.45
	Set 1	85	2.54–24.60	6.91	2.61	37.77
	Set 2	15	2.54–9.53	6.42	2.61	40.65
Apigenin-7-glucoside	Set 1 + 2	100	1.93–2.47	2.04	0.35	17.10
	Set 1	85	1.93–2.47	2.12	0.35	16.50
	Set 2	15	1.94–2.20	2.11	0.11	5.21

**Table 1.** Statistics of the samples sets for phenolic compounds quantification in *Olea europaea L.*, leaves. *SD*, standard deviation; *CV*, coefficient of variation calculated as *SD/mean*.

levels of 0.29–0.24 mg/g, respectively<sup>25</sup>. However, Da Silva et al.<sup>26</sup> showed similar values to those obtained. The hydroxytyrosol levels (Table 1) were lower than those obtained in the research of Ansari et al.<sup>21</sup>, whereas compared to those obtained by Martínez-Navarro et al.<sup>23</sup> and Da Silva et al.<sup>26</sup>, the concentration values were within the same range. For apigenin-7-glucoside, it also appeared at a lower concentration compared to Ansari et al.<sup>21</sup> research. Furthermore, the standard deviation found for this compound, compared to the other compounds analysed, was low due it remained in a close range.

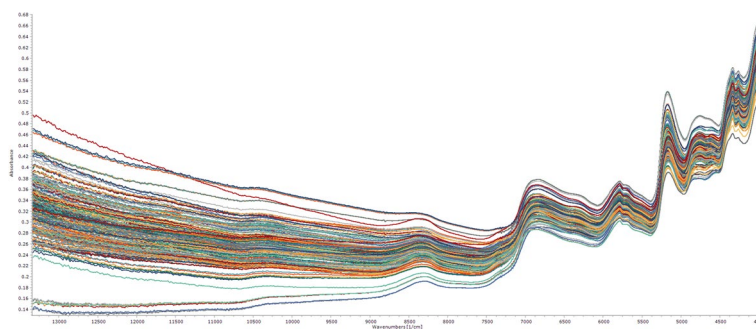
In relation to the coefficients of variation (CV) for the above-mentioned compounds, the values range from 17.10 to 73.97% when the entire range was considered, which shows that, although the olive tree variety used was the same, there was a large variability in the analyte's levels between different olive trees, the cultivation method and its origin, an aspect previously observed in the bibliography. Thus, Kabbash et al.<sup>27</sup> showed differences between levels of oleuropein in leaves from different origins. Also, factors such as temperature and water stress affect the levels of these compounds, giving differences in their concentrations depending on the origin plantation and the conditions experienced<sup>28</sup>. Along with the previous, it was demonstrated differences in the phenolic metabolite profile during the collecting time of the year<sup>23,29</sup>.

### NIR spectra interpretation for *Olea europaea L.* leaves metabolites

The possibility of analyzing *Olea europaea L.* leaf metabolites using near-infrared spectroscopy (NIR) offers several advantages over high-performance liquid chromatography (HPLC), which is the usual technique to determine the phenolic composition of the olive leaves<sup>13,27,30</sup>. Among the advantages to highlight, the following are included: faster and more efficient analysis, with real-time measurements and no need for extensive sample preparation. In fact, NIR is a non-destructive technique, allowing samples to be reused, and it is easier to operate and automate. Additionally, it significantly reduces operational costs by not requiring solvents, reagents, or expensive consumables, and it has lower maintenance requirements. Therefore, the use of NIR for the determination of olive leaf metabolites offers an advantage in many aspects compared to the determination by HPLC–DAD. The range used for scanning samples was 13333.33–4000  $\text{cm}^{-1}$  (750–2500 nm), after which the range was adjusted for each calibration model to improve the method.

The bibliography related to NIR spectroscopy used in olive leaves samples is limited<sup>9</sup>. However, it is possible to associate the measurement ranges of the spectra (Fig. 1) with different types of chemical bonds. In this way, bands corresponding to bonds occurring in phenolic compounds can be detected in the spectrum, since they are widely studied in the literature<sup>31,32</sup>.

As it is well known, phenolic compounds consist of one or more aromatic rings and at least one hydroxyl group directly linked to the rings<sup>33</sup>, which present common bands associated with these structures. Six-membered aromatic rings show two or three bands due to C=C ring stretching vibrations in the region of 1625–1430  $\text{cm}^{-1}$ <sup>34</sup>. Also, aromatic C=C stretching vibrations of varying intensity and position can be observed around 1625–1590  $\text{cm}^{-1}$ , 1590–1575  $\text{cm}^{-1}$ , 1525–1470  $\text{cm}^{-1}$ , and 1465–1430  $\text{cm}^{-1}$ <sup>34</sup>, depending on the number of ring substituents and their positions. In phenols, the interaction of O–H deformation and C=O stretching vibrations can occur in the spectral regions of 1390–1330  $\text{cm}^{-1}$  and 1260–1180  $\text{cm}^{-1}$ , being the bands corresponding to COH deformation appear in the region of 1382–1317  $\text{cm}^{-1}$ <sup>34</sup>. The peaks near 8695  $\text{cm}^{-1}$  could be interpreted as a blend of symmetric and asymmetric O–H stretching bands along with O–H bending<sup>35</sup>. Variations in the absorption peaks near 13333.33  $\text{cm}^{-1}$  and 11764.71  $\text{cm}^{-1}$  could be attributed to the third overtone of the H<sub>2</sub>O and C–H functional groups, respectively<sup>36</sup>.



**Fig. 1.** Near-infrared spectrum of *Olea europaea L.*, leaves analysed.

Phenolic compound	Spectral range (cm <sup>-1</sup> )	Calibration model	Pre-process	Number of samples	R <sup>2</sup> <sub>cv</sub>	RMSEC	RMSECV	RMSEP	RPD <sub>c</sub>	RPD	RPD <sub>p</sub>
Oleuropein	11111.11 – 4000	CVP1	MSC – 1.5.9	Set 1 + 2	0.93	2.41	4.52	9.65	4.29	9.72	2.78
		CVP2		Set 1	0.93	25.85	11.52		0.96	3.82	
				Set 2				11.51			2.22
Hydroxytyrosol	11111.11 – 4000	CVP1	MSC – 1.9.13	Set 1 + 2	0.93	0.16	0.16	0.15	4.58	4.89	2.80
		CVP2		Set 1							
				Set 2							
Hydroxytyrosol hexoside	13333.33 – 4000	CVP1	MSC – 1.9.5	Set 1 + 2	0.95	0.00	0.18	0.18	4.7	9.79	1.73
		CVP2		Set 1	0.90	0.02	0.27		2.09	6.62	
				Set 2				0.27			1.73
Verbascoside	13,333.33 – 4000	CVP1	MSC – 1.9.5	Set 1 + 2	0.90	0.00002	0.9	0.89	3.91	4.20	1.40
		CVP2		Set 1	0.90	0.38	0.95		1.46	3.94	
				Set 2				0.95			1.44
Apigenin-7-glucoside	11111.11 – 4000	CVP1	MSC – 1.5.5	Set 1 + 2	0.96	0.00	0.01	0.04	2.60	2.89	8.00
		CVP2		Set 1	0.85	0.36	0.94		1.44	4.02	
				Set 2				0.93			1.48

**Table 2.** NIR statistical parameters for the calibration, cross-validation and prediction data of the measured *Olea europaea L.*, leaves phenolic compounds. CVP, Calibration Validation Prediction; MSC, Multiplicative Scatter Correction; preprocessing applied to correct for light scattering in the samples. The three-digit code (a.b.c.) where ‘a’ refers to the number of the derivative (1 and 2), ‘b’ is the interval over which the derivative is calculated (5, 9, and 13), ‘c’ corresponds to the number of data points in a smoothing (5, 13, and 9). R<sup>2</sup><sub>cv</sub>: coefficient of determination in cross validation; RMSEC, root mean squared error of calibration; RMSECV, root mean squared error of cross validation; RMSEP, root mean squared error of prediction; RPD<sub>c</sub>, residual predictive deviation (SD/SECv); RPD: calculated according to Yılmaz-Düzüyanan et al.<sup>15</sup>; RPD<sub>p</sub>: SD<sub>cal</sub>/SEP.

### NIR calibration and validation

The more accurate model, for each analyzed compound and calibration model, obtained with NIR raw spectral data regressed against their HPLC–DAD determination is summarized in Table 2. In it, the statistics of the prediction models for cross-validation of the calibrations and of the test validation of the compounds are summarized in Table 1, along with selected spectral range, applied pre-processing and number of samples for both calibration models CVP1 and CVP2. For calibration model 1 (CVP1), which encompasses Set 1 + 2, good calibrations, validations and predictions were obtained for all compounds. However, for calibration model 2 (CVP2), through which it was calibrated and validated with one set (Set 1) and predicted with other (Set 2), a good calibration was achieved for all the phenolic compounds, except for hydroxytyrosol, for which it was not possible to obtain a good calibration model.

Despite to the fact that the values of the bands related to the target compounds of this study are known<sup>34</sup>, and have been previously indicated, for the calibration model’s development, the entire infrared spectral region acquired was considered, although with some differences depending on the calibrated compound (Table 2). For oleuropein, the spectral range used was similar that used in other research for oleuropein calibration in olive oils, where the selected range was 10,000–4000 cm<sup>-19</sup>. Although there is no literature that specifies a specific band for the other phenolic compounds studied (hydroxytyrosol, verbascoside, hydroxytyrosol hexoside and apigenin-7-glucoside) in olive leaves, their calibration in olive oil has been carried out according to the total phenolic compounds range, using a range of 12,500–4000 cm<sup>-137</sup>.

Regarding the spectral processing method, and to develop a good prediction model and obtain a high R<sub>c</sub><sup>2</sup> and R<sub>cv</sub><sup>2</sup>, it was necessary to apply the first derivative to the spectra of the samples for all the compounds to

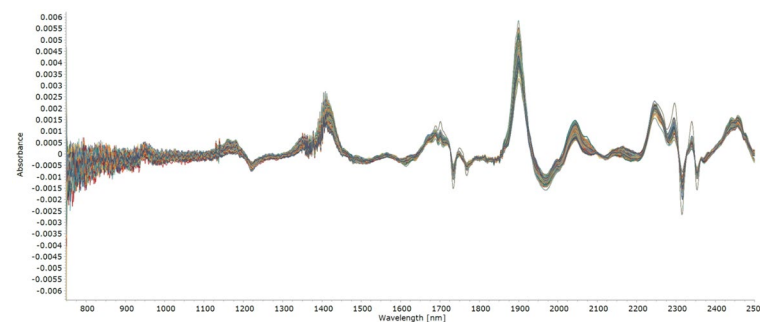
be calibrated. The first derivative of NIR spectra has the effect of highlighting spectral features by eliminating the baseline and reducing background variations. This facilitates the identification and comparison of peaks, providing better resolution of changes in the concentration of components present in the samples. After applying the first derivative to the spectrum, it became possible to differentiate the calibration of the compounds at 9 distinct interval points for hydroxytyrosol, hydroxytyrosol hexoside, and verbascoside, and at 5 interval points for oleuropein and apigenin-7-glucoside. Figure 2 shows the first derivative of all measured spectra, where it can be seen how by applying the derivative, the variations in the signal obtain greater definition compared to the untreated spectrum (Fig. 1). The largest differences between the spectra were observed in the peak at  $8500\text{ cm}^{-1}$ , which has been related to the differences in the mineral composition of the leaves<sup>38</sup>. Another notable peak at  $5798\text{ cm}^{-1}$  corresponds to the primary vibration of the C–H group, commonly associated with lipids, oils, and aromatic substances<sup>39</sup>. The spectra range of  $7500\text{--}6800\text{ cm}^{-1}$  and  $5500\text{--}4600\text{ cm}^{-1}$  were selected in other works to perform calibrations of phenolic compounds<sup>37</sup>. However, when using these ranges, good results were not obtained. So, to perform the calibration of individual phenolic compounds in leaves, it was necessary to resort to a wider range ( $13333.33\text{--}4000\text{ cm}^{-1}$ ) to obtain good values for the calibration indices. As for smoothing, 5 points were added for hydroxytyrosol hexoside, verbascoside and apigenin-7-glucoside, whereas for oleuropein, 9 smoothing points were necessary and 13 points for hydroxytyrosol.

PLS-R was used to perform the calibration model, increasing the performance of the predictive models in the selected spectral range by reducing dimensionality and improving the accuracy of predictions as well as the ease of interpreting the components. After the calibration, for both models (CVP1 and CVP2), the validation was carried out by means of cross-validation, where the  $R^2_{cv}$  values of each validation model were obtained (Table 2). The  $R^2_{cv}$  value reflects the goodness of the predictive model that fits the validation data, representing the proportion of the variance in the responses that can be explained by the predictor variables. The range of the  $R^2_{cv}$  values could vary from 0 to 1, the latter being the best  $R^2_{cv}$  value that can be obtained, indicating a high predictive power of the model. Comparing the  $R^2_{cv}$  values for both calibration models, the highest values were observed for CVP1. For CVP2, good model predictions were also obtained for all compounds as the values, although lower, were quite high (0.9), being the same value for oleuropein (0.93) and slightly lower in the case of apigenin-7-glucoside, with a  $R^2_{cv}$  value of 0.85. Contrary, with this last calibration model, for hydroxytyrosol, no good calibration values were obtained, so the validation and prediction could not be obtained. This may be due to the low concentration of the compound in the olive leaf samples which leads to a low dispersion of the data.

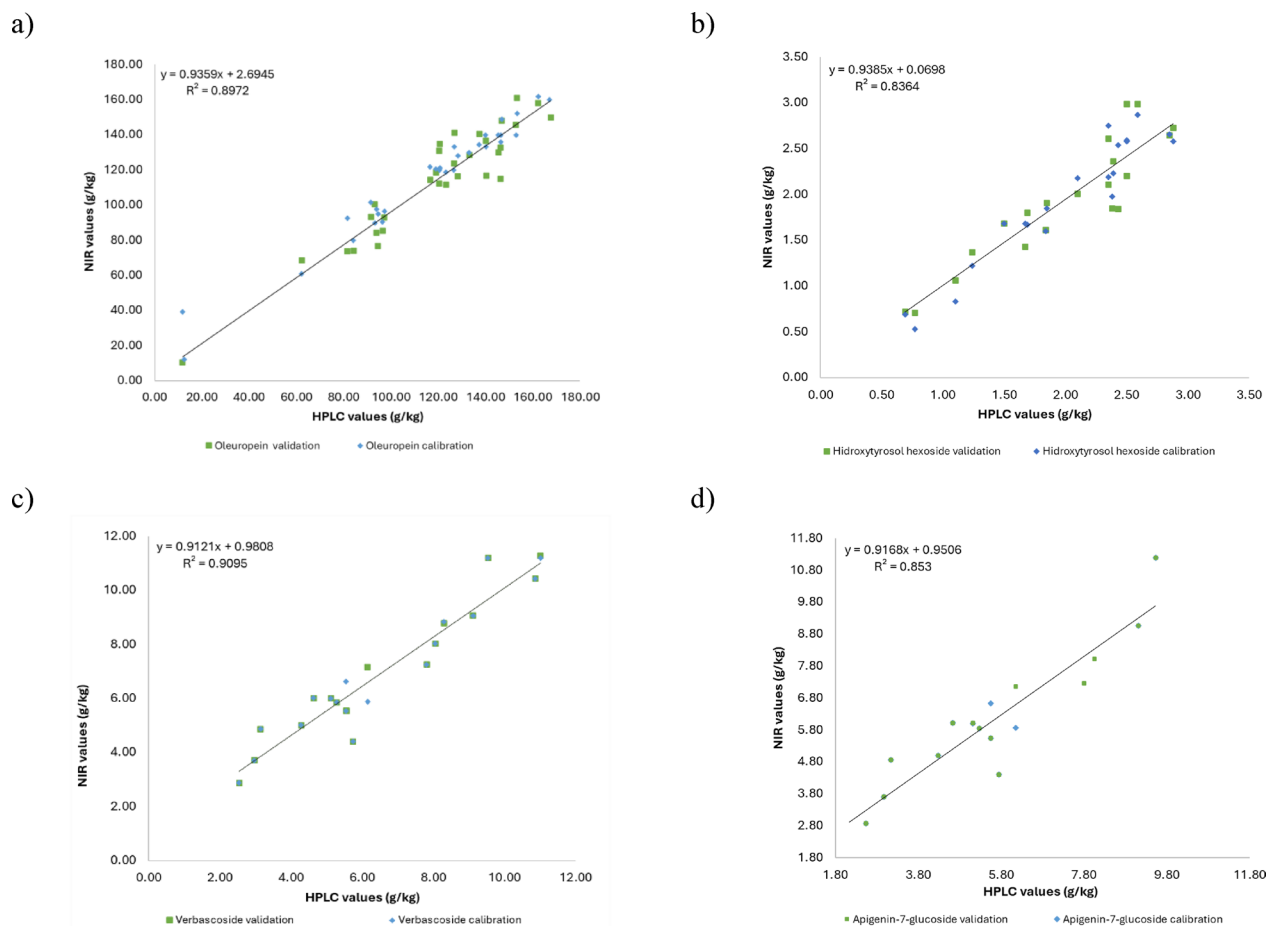
Along with the  $R^2_{cv}$  values obtained for the different calibrated compounds, Table 2 shows the indices that inform about predictive ability of the method. The RPD index (Ratio of Performance to Deviation) is a key metric used to evaluate the accuracy of a model, so, is critical for determining the model's applicability. It is considered that high RPD ( $RPD > 2.5$ ) indicates high accuracy, suitable for precise quantitative predictions; a medium RPD ( $1.5 < RPD \leq 2.5$ ) suggests moderate accuracy, useful for classification; and, a low RPD ( $RPD \leq 1.5$ ) reflects low accuracy and predictive limitations, useful only for large differences in the data<sup>40</sup>. The values calculated according to Luo et al.<sup>18</sup> in CVP2 for oleuropein was very close to 4 as for the rest of the compounds studied, except for hydroxytyrosol hexoside, where the RPD was 6.62, indicating that the calibration model has good predictive capability<sup>40</sup>. The RPD values obtained for CVP2 were lower than those obtained for CVP1, so CVP2 could have had a worse ability to predict samples than CVP1 for all the compounds studied. Also, CVP2 has a higher RMSEP than CVP1 which is indicative of an overfitting of the model, giving a worse predictive ability when external samples are used. These results are since for the CVP1 model the same samples were used for calibration, validation and prediction but for the CVP2 model the samples were divided into Set 1 and Set 2 as shown in Table 2.

Verbasoside in the CVP2 model obtained an RPD close to 4 (good predictive ability) and an  $RPD_c$  close to 2, which could be considered a good result, with a higher RMSEP than in CVP1 indicating that the CVP1 model has more overfitting. In CVP1 the RPD and  $RPD_c$  were 4.20 and 3.91, which could indicate good predictive ability and the RMSEP of CVP1 was lower than the RMSEP of CVP2. In relation to their  $RPD_p$  value, both CVP1 and CVP2 were the lowest obtained due to their low deviation. The RMSEC values were low, which is also indicative of the good accuracy of the selected model. In CVP1, a good calibration of the hydroxytyrosol was achieved, in contrast to CVP2, which could be due to the presence of a larger number of samples in the model.

In the case of apigenin-7-glucoside and hydroxytyrosol hexoside for the CVP1 model, RMSEC value was 0 due to the small difference between the predicted values and those measured by HPLC–DAD. The number



**Fig. 2.** First derivative of all *Olea europaea* L., leaves near-infrared spectra.



**Fig. 3.** Scatter plot obtained with HPLC–DAD versus values predicted by the NIR in *Olea europaea* L., leaves using PLS-R in CVP2: (a) Oleuropein; (b) Hydroxytyrosol hexoside; (c) Verbascoside; (d) Apigenin-7-glucoside.

of samples containing apigenin-7-glucoside were lower than for other compounds, leading to a lower RMSEP value, which reduces the predictive ability with external samples due to over fitting of the model. As for the results obtained in CVP2, the RMSEC values were higher and the RPD in the case of apigenin-7-glucoside, but lower for hydroxytyrosol hexoside.

When comparing both models, CVP1 shows better  $R_{cv}^2$  and RPD (except in the case of apigenin-7-glucoside, where the  $RPD_c$  of CVP2 was lower than CVP1) which should indicate that CVP1 could be more appropriate for measuring such compound in olive leaf samples than CVP2. However, given that the CVP1 model has been developed with the same samples for both calibration and validation/prediction, it could be considered as less appropriate for measuring samples outside the calibration process as it is more likely to over-fit the samples. For this reason, Fig. 3 shows scatter plots of the values of olive leaves compounds predicted by NIR versus the values obtained with HPLC–DAD using the CVP2 calibration model. Good correlations, except for hydroxytyrosol (data not showed) were observed, indicating that the values obtained by both methods (HPLC–DAD and NIR) were very similar. The best correlations were obtained for verbascoside, yielding  $R^2$  values of 0.91, followed by oleuropein (0.89), being the values for hydroxytyrosol hexoside and apigenin-7-glucoside 0.83 and 0.85, respectively. The differences observed for the  $R^2$  values are justified by the high concentration of oleuropein and verbascoside in the Cornicabra olive leaves, which facilitated the achievement of good predictive values.

## Conclusions

The development of a calibration method using Near-Infrared Spectroscopy (NIR) presents a promising tool to determine the principal phenolic compounds in *Olea europaea* L. leaves: oleuropein, verbascoside, hydroxytyrosol hexoside and apigenin-7-glucoside. In the case of hydroxytyrosol, it was only possible to obtain a calibration for one of the tested models. The CVP2 model proved to be more reliable for the prediction of external samples since different samples were used for each step of the process. Although CVP1 shows good rates, using the same samples to calibrate-validate -predict leads to a lower ability to predict external samples as well as a higher overfitting of the model.

In CVP2 model, the detailed data pre-processing methods and the adjustment of wavenumber ranges resulted in high predictability and the precise estimation of the mentioned metabolites except for hydroxytyrosol, for

which a good calibration has not been obtained. In particular, the oleuropein calibration was the most successful, with the best predictive power, which could be related to the high concentration in olive leaves. For verbascoside and hydroxytyrosol hexoside, the predictive models showed acceptable predictive capabilities. The calibration of hydroxytyrosol did not give good results due to the low presence of the compound in the samples selected for calibration and the low number of samples.

The determination of these high added value compounds in the olive leaf by precise and rapid methods is crucial to efficiently harness the potential of the olive leaves.

### Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 11 March 2025; Accepted: 30 July 2025

Published online: 14 August 2025

### References

- Bilal, R. M. et al. Olive oil: Nutritional applications, beneficial health aspects and its prospective application in poultry production. *Front. Pharmacol.* **12**, 723040 (2021).
- Espeso, J. et al. Olive leaf waste management. *Front. Sustain. Food Syst.* **5**, 660582 (2021).
- Guinda, Á. et al. Determination of major bioactive compounds from olive leaf. *LWT Food Sci. Technol.* **64**, 431–438 (2015).
- Lucas, L., Russell, A. & Keast, R. Molecular mechanisms of inflammation. Anti-inflammatory benefits of virgin olive oil and the phenolic compound oleocanthal. *Curr. Pharm. Des.* **17**, 754–768 (2011).
- Vijakumaran, U. et al. Effects of hydroxytyrosol in endothelial functioning: A comprehensive review. *Molecules* **28**, 1861 (2023).
- Nediani, C., Ruzzolini, J., Romani, A. & Calorini, L. Oleuropein, a bioactive compound from *Olea europaea* L., as a potential preventive and therapeutic agent in non-communicable diseases. *Antioxidants* **8**, 578 (2019).
- Bernini, R., Merendino, N., Romani, A. & Velotti, F. Naturally occurring hydroxytyrosol: Synthesis and anticancer potential. *Curr. Med. Chem.* **20**, 655–670 (2013).
- Robles-Almazan, M. et al. Hydroxytyrosol: Bioavailability, toxicity, and clinical applications. *Food Res. Int.* **105**, 654–667 (2018).
- Can, A. et al. The potential of near and mid-infrared spectroscopy for rapid quantification of oleuropein, total phenolics, total flavonoids and antioxidant activity in olive tree (*Olea europaea* L.) leaves. *J. Food Meas. Charact.* **12**, 2747–2757 (2018).
- Comino, F., Ayora-Cañada, M. J., Aranda, V., Díaz, A. & Domínguez-Vidal, A. Near-infrared spectroscopy and X-ray fluorescence data fusion for olive leaf analysis and crop nutritional status determination. *Talanta* **188**, 676–684 (2018).
- Dupuy, N., Galtier, O., Ollivier, D., Vanloot, P. & Artaud, J. Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model application to virgin olive oil analysis. *Anal. Chim. Acta.* **666**, 23–31 (2010).
- García Martín, J. F. Potential of near-infrared spectroscopy for the determination of olive oil quality. *Sensors* **22**, 2831 (2022).
- Martínez-Navarro, et al. Development and validation of an HPLC-DAD method for determination of oleuropein and other bioactive compounds in olive leaf by-products. *J. Sci. Food Agric.* **101**, 1447–1453 (2021).
- Isaksson, T. & Naes, T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl. Spectrosc.* **42**, 1273–1284 (1988).
- Yılmaz-Düzyaman, H., De La Rosa, R., Velasco, L., Núñez-Sánchez, N. & León, L. Oil quality prediction in olive oil by near-infrared spectroscopy: Applications in olive breeding. *Agriculture* **14**, 721 (2024).
- Valinger, D. et al. Development of near infrared spectroscopy models for quantitative prediction of the content of bioactive compounds in olive leaves. *Chem. Biochem. Eng. Q.* **32**, 535–543 (2018).
- Sileoni, V., Van Den Berg, F., Marconi, O., Perretti, G. & Fantozzi, P. Internal and external validation strategies for the evaluation of long-term effects in NIR calibration models. *J. Agric. Food Chem.* **59**, 1541–1547 (2011).
- Luo, Z., Thorp, K. R. & Abdel-Haleem, H. A high-throughput quantification of resin and rubber contents in *Parthenium argentatum* using near-infrared (NIR) spectroscopy. *Plant Methods* **15**, 154 (2019).
- Suchat, S. et al. Fast determination of the resin and rubber content in *Parthenium argentatum* biomass using near infrared spectroscopy. *Ind. Crops Prod.* **45**, 44–51 (2013).
- Anjos, O. et al. PLS-R calibration models for wine spirit volatile phenols prediction by near-infrared spectroscopy. *Sensors* **22**, 286 (2022).
- Ansari, M., Kazempour, M. & Fathi, S. Iranian chemical society development of a simple green extraction procedure and HPLC method for determination of oleuropein in olive leaf extract applied to a multi-source comparative study. *J. Iran. Chem. Soc.* **8**, 38–47 (2011).
- Ortega-García, F. & Peragón, J. HPLC analysis of oleuropein, hydroxytyrosol, and tyrosol in stems and roots of *Olea europaea* L. cv. picual during ripening. *J. Sci. Food Agric.* **90**, 2295–2300 (2010).
- Martínez-Navarro, M. E., Cebrián-Tarancón, C., Salinas, M. R. & Alonso, G. L. Evolution of oleuropein and other bioactive compounds in arbequina olive leaves under different agronomic conditions. *Horticulturae* **8**, 530 (2022).
- Talhaoui, N. et al. Determination of phenolic compounds of 'Sikitita' olive leaves by HPLC-DAD-TOF-MS. Comparison with its parents 'Arbequina' and 'Picual' olive leaves. *LWT* **58**, 28–34 (2014).
- Martínez-Navarro, M. E. et al. Quantitative determination of the main phenolic compounds, antioxidant activity, and toxicity of aqueous extracts of olive leaves of Greek and Spanish genotypes. *Horticulturae* **9**, 55 (2023).
- da Silva, P. S. et al. Determination of an optimum extraction region for the recovery of bioactive compounds from olive leaves (*Olea europaea* L.) using green dynamic pressurized liquid extraction. *Braz. J. Chem. Eng.* **40**, 843–861 (2023).
- Kabbash, E. M., Abdel-Shakour, Z. T., El-Ahmady, S. H., Wink, M. & Ayoub, I. M. Comparative metabolic profiling of olive leaf extracts from twelve different cultivars collected in both fruiting and flowering seasons. *Sci. Rep.* **13**, 612 (2023).
- Mechri, B., Tekaya, M., Hammami, M. & Chehab, H. Effects of drought stress on phenolic accumulation in greenhouse-grown olive trees (*Olea europaea* L.). *Biochem. Syst. Ecol.* **92**, 104112 (2020).
- Difonzo, G. et al. Metabolomics approach to characterize green olive leaf extracts classified based on variety and season. *Plants* **11**, 3321 (2022).
- Ghomari, O. et al. Phenolic profile (HPLC-UV) of olive leaves according to extraction procedure and assessment of antibacterial activity. *Biotechnol. Rep.* **23**, e00347 (2019).
- Cascant, M. M. et al. Determination of total phenolic compounds in compost by infrared spectroscopy. *Talanta* **153**, 360–365 (2016).
- Kljusuric, J. G., Mihalev, K. & Becic, I. Near-infrared spectroscopic analysis of total phenolic content and antioxidant activity of berry fruits. *Food Technol. Biotechnol.* **54**, 236–242 (2016).
- Luna-Guevara, M. L., Luna-Guevara, J. J., Hernández-Carranza, P., Ruiz-Espinosa, H. & Ochoa-Velasco, C. E. Phenolic compounds: A good choice against chronic degenerative diseases. *Stud. Nat. Prod. Chem.* **59**, 79–108 (2018).

34. Abbas, O. et al. Phenolic compound explorer: A mid-infrared spectroscopy database. *Vib. Spectrosc.* **92**, 111–118 (2017).
35. Bellincontro, A. et al. Feasible application of a portable NIR-AOTF tool for on-field prediction of phenolic compounds during the ripening of olives for oil production. *J. Agric. Food Chem.* **60**, 2665–2673 (2012).
36. Xia, Y., Huang, W., Fan, S., Li, J. & Chen, L. Effect of spectral measurement orientation on online prediction of soluble solids content of apple using Vis/NIR diffuse reflectance. *Infrared Phys. Technol.* **97**, 467–477 (2019).
37. Mora-Ruiz, M. E. et al. Assessment of polar phenolic compounds of virgin olive oil by NIR and mid-IR spectroscopy and their impact on quality. *Eur. J. Lipid Sci. Technol.* **119**, 1600099 (2017).
38. Fernández, V. M., Fernández-Cabanas, E., Garrido-Varo, A., Delgado-Pertinẽez, M. & Gómez-Cabrera, A. Nutritive evaluation of olive tree leaves by near-infrared spectroscopy: Effect of soil contamination and correction with spectral pretreatments. *Appl. Spectrosc.* **62**, 51–58 (2008).
39. Frizon, C. N. T. et al. Determination of total phenolic compounds in yerba mate (*Ilex paraguariensis*) combining near infrared spectroscopy (NIR) and multivariate analysis. *LWT–Food Sci. Technol.* **60**, 795–801 (2015).
40. Prieto, N., Pawluczyk, O., Dugan, M. E. R. & Aalhus, J. L. A review of the principles and applications of near-infrared spectroscopy to characterize meat, fat, and meat products. *Appl. Spectrosc.* **71**, 1403–1426 (2017).

## Acknowledgements

Authors wish to thank the Ministry of Education, Universities and Research of the Community Board of Castilla-La Mancha, and the European Regional Development Fund (FEDER) for financing this work through the AZUVOL II project (ref.: SBPLY/21/180225/000002 and the Universidad de Castilla-La Mancha for the financed Project 2023-GRIN-34180. Authors also wish to thank the agreements for which the D.O. for Extra Virgin Olive Oil (Montes de Toledo, Campo de Calatrava and Campo de Montiel), and to the “Marca Colectiva Gráfica de la Asociación Mesa del Aceite Sierra de Alcaraz” (Castilla-La Mancha, Spain) and the farmers have supplied the samples and supported this research.

## Author contributions

Jorge Francisco Escobar Talavera did the experiments and wrote the manuscript. María Esther Martínez-Navarro sample analysis Gonzalo L Alonso Diaz-Marta-director and reviewer Rosario Sánchez Gómez-director and reviewer.

## Declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-14292-6>.

**Correspondence** and requests for materials should be addressed to R.S.-G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025