

Barbara Muik · Bernhard Lendl · Antonio Molina-Díaz  
Luis Pérez-Villarejo · María José Ayora-Cañada

## Determination of oil and water content in olive pomace using near infrared and Raman spectrometry. A comparative study

Received: 10 October 2003 / Revised: 16 December 2003 / Accepted: 23 December 2003 / Published online: 17 February 2004  
© Springer-Verlag 2004

**Abstract** Near infrared (NIR) reflectance and Raman spectrometry were compared for determination of the oil and water content of olive pomace, a by-product in olive oil production. To enable comparison of the spectral techniques the same sample sets were used for calibration (1.74–3.93% oil, 48.3–67.0% water) and for validation (1.77–3.74% oil, 50.0–64.5% water). Several partial least squares (PLS) regression models were optimized by cross-validation with cancellation groups, including different spectral pretreatments for each technique. Best models were achieved with first-derivative spectra for both oil and water content. Prediction results for an independent validation set were similar for both techniques. The values of root mean square error of prediction (RMSEP) were 0.19 and 0.20–0.21 for oil content and 2.0 and 1.8 for water content, using Raman and NIR, respectively. The possibility of improving these results by combining the information of both techniques was also tested. The best models constructed using the appended spectra resulted in slightly better performance for oil content (RMSEP 0.17) but no improvement for water content.

**Keywords** Raman · Near Infrared · Oil content · Water content · Olive pomace · Partial least squares (PLS) regression

### Introduction

Olive pomace is a by-product of the mechanical processing of olives in production of virgin olive oil. The process

of releasing the oil from the plant tissue begins by milling the olives to tear the flesh cells in order to let the oil run out of the vacuoles. This is followed by stirring the olive paste to permit the formation of large drops of oil and to break up the oil–water emulsion. In so-called “dual-phase decanters” the oil is then separated by direct continuous centrifugation from the pomace, which consists of vegetable matter and water. The yield of oil varies from 80 to 90% of the total oil content of the olives, because the oil in the olive paste is only partially free to escape and part of it remains in the unbroken cells or is trapped in the tissues of the cytoplasm, or is emulsified in the aqueous phase [1]. Determination of residual oil in the olive pomace is regarded as crucial control aspect of quantitative optimization of the olive oil production plant. Furthermore, the olive pomace is sold to other factories, where recovery of residual oil is carried out by solvent extraction to obtain the so-called “olive pomace oil”. For this, the olive pomace must be dried, hence oil and water content determine its economic value.

The need for process control demands analytical methods capable of rapid analysis of the olive pomace to enable adjustment of appropriate operating conditions. Spectroscopic methods are well suited for process-control applications because they are fast and can be easily implemented in-line or at-line, giving better control of the entire production [2]. Among spectroscopic techniques, near infrared (NIR) and Raman spectrometry have the advantage that they provide direct molecular specific information and do not require extensive sample preparation. In combination with chemometrics these techniques can successfully replace many time-consuming chemical methods. Thus, for analysis of olive pomace these techniques for determination of oil content can be convenient alternatives to the traditional Soxhlet method [3], which requires a time-consuming drying step followed by solvent extraction. This method has already been substituted in many routine laboratories by nuclear magnetic resonance (NMR) spectroscopy. But, even the use of this technique lacks rapidness, because water interferes strongly and, therefore, the olive pomace sample must be completely dry. Consequently, this

B. Muik · A. Molina-Díaz · L. Pérez-Villarejo  
M. J. Ayora-Cañada (✉)  
Department of Physical and Analytical Chemistry,  
University of Jaén, 23071 Jaén, Spain  
e-mail: mjayora@ujaen.es

B. Muik · B. Lendl  
Institute of Chemical Technologies and Analytics,  
Vienna University of Technology,  
Getreidemarkt 9/164, 1060 Vienna, Austria

method is unsuitable for process control. In contrast, both NIR and Raman spectroscopy can extract useful analytical information about oil and water content working directly on the olive pomace without any sample pretreatment.

In recent years NIR spectrometry has frequently been applied in food industry. NIR spectroscopic analysis of dairy products such as milk [4, 5], meat [6], butter [7], and others [8] have been reported. In contrast, few examples of application of Raman spectroscopy to quantitative analysis and quality control in the food industry have been reported [9, 10]. In previous work we demonstrated the suitability of Fourier transform Raman spectrometry for determination of the water and oil content of olives [11], and the expected acidity of the oil produced from the analyzed olives [12]. Here we approach the analysis of olive pomace using this technique and compare the performance of this method with a method based on NIR diffuse reflectance spectrometry. Analysis of olive pomace using an NIR instrument equipped with optical filters has previously been reported. Measurements at four wavenumbers were used to develop a calibration by multiple linear regression (MLR) [13]. In this work, we compare the performance of NIR diffuse reflectance and Raman spectrometry using Fourier transform spectrometers and partial least squares (PLS) regression for analysis of olive pomace.

## Experimental

### Instrumentation

Fourier transform near infrared spectra were measured in diffuse reflectance mode using an Antaris near-IR analyzer (Thermo Nicolet Corporation). The instrument is equipped with an integrating sphere that contains an internal gold reference.

Fourier transform Raman spectra were obtained by use of an RFS 100/S spectrometer (Bruker Optics). The 1064 nm line from a Nd:YAG laser (Coherent) was used for excitation.

### Software

TQ Analyst 6.1.1 (Thermo Nicolet Corp.) was used for all PLS models as well as for all preprocessing techniques. NIR spectra were used directly in the spectral range from 4000 to 10000  $\text{cm}^{-1}$ . Raman spectra in the spectral range from 20 to 3800  $\text{cm}^{-1}$  were imported.

For combined models the spectra were imported to Microsoft Excel 2000. With this program the two spectra of each sample were combined into one.

### Samples and reference analysis

A total of 132 olive pomace samples were obtained from different oil-production plants in the province of Jaén (Spain) in the period between December 2002 and February 2003.

Reference analyses were performed at CM Europa SA laboratory, Martos, (Spain). To determine water content, the sample (45 g) was dried for 12 h in an oven at 105 °C. The loss of weight gave the amount (%) of water and volatile matter in the sample [14]. The dried sample was then used to measure the oil content, by NMR spectroscopy. Soxhlet extraction was used for calibration and for quality control. For this the oil was extracted with hexane for 6 h using the Soxhlet extractor. After evaporation of the solvent the oil content was determined gravimetrically.

## Measurement procedures

### Near infrared

Olive pomace samples were placed in a sample cup (diameter 3 cm), which was placed on top of the integrating sphere optics and rotated at 10 revolutions per minute during measurement. All spectra were obtained at a resolution of 8  $\text{cm}^{-1}$  and averaged over 50 scans; this resulted in an acquisition time of 1 min.

### Raman

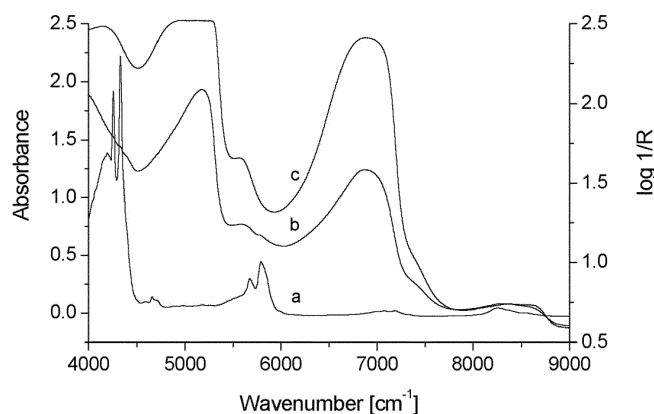
The samples were investigated in a home-made cell comprising a hollow cylindrical magnet covered inside with a Teflon layer. The dimensions of the cell were 13 mm inner diameter and 5 mm depth. The cell was attached to a synchronous motor mounted in the sample compartment. The motor enabled eccentric rotation of the sample cell around the horizontal axis of the laser beam at five revolutions per minute. All spectra were obtained with 500-mW laser power, at a resolution of 8  $\text{cm}^{-1}$  and were averages from 400 scans, resulting in an acquisition time of 5 min.

## Results and discussion

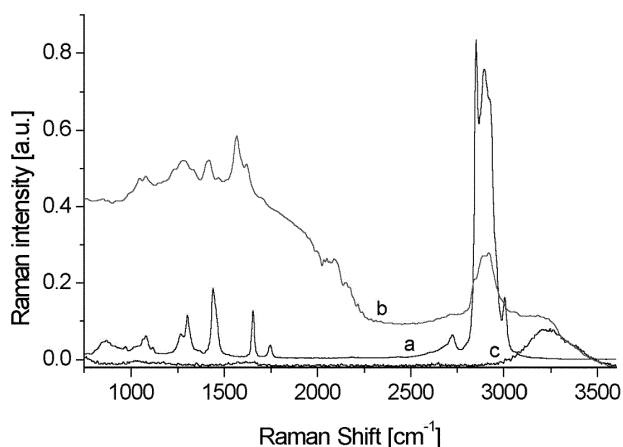
### Spectral features

#### Near Infrared

Figure 1 shows the near Infrared spectra of olive pomace, oil, and water. Two broad bands at 7500–6150  $\text{cm}^{-1}$  and 5350–4550  $\text{cm}^{-1}$  dominate the spectrum of the olive pomace. These bands are assigned to the first overtone of the –OH stretching vibration and to combination bands of the water molecules, respectively. The significant bands from the pure olive oil are also visible, but far less intense. The most intense bands in the oil spectra can be found at 4260 and 4370  $\text{cm}^{-1}$ , which are characteristic of the combination of CH-stretching vibrations of –CH<sub>3</sub> and –CH<sub>2</sub> with other vibrations. The two bands at 5700 and 5750  $\text{cm}^{-1}$  correspond to the first overtone of the C–H stretching vibration of –CH<sub>3</sub>, –CH<sub>2</sub>, and –HC=CH–. In the region between 7700 and 9100  $\text{cm}^{-1}$  the second overtone of the C–H stretching vibration of –CH<sub>3</sub>, –CH<sub>2</sub> and –HC=CH– can be found [15].



**Fig. 1** NIR spectra of (a) oil (transmission, pathlength 5 mm), (b) olive pomace (diffuse reflectance), and (c) water (transmission, pathlength 0.5 mm)



**Fig. 2** Raman spectra of (a) oil, (b) olive pomace (intensities amplified by a factor of 6), and (c) water (intensities amplified by a factor of 10)

### Raman

The FT-Raman spectra of olive pomace, oil and water are presented in Fig. 2. The major bands of the virgin olive oil can be seen at 1267 (in-plane  $\delta(\text{C-H})$  deformation in unconjugated *cis* double bond) 1302 (in-phase methylene twisting motion), 1442 ( $\delta(\text{CH}_2)$ ), 1655 ( $\nu(\text{C=C})$ ), 1747  $\text{cm}^{-1}$  ( $\nu(\text{C=O})$ ), 2852 ( $\nu_{\text{sym}}(\text{CH}_2)$ ), and 2900  $\text{cm}^{-1}$  ( $\nu_{\text{sym}}(\text{CH}_3)$ ) [16]. A high fluorescence background characterizes the spectra of olive pomace. The spectral features of the vegetable matter, especially the olive kernel, overlap with the oil bands. The most characteristic band is located at 1604  $\text{cm}^{-1}$  and can be assigned to the aromatic ring stretch of lignin, a major component of the olive kernel. The broad band centered at 3250  $\text{cm}^{-1}$ , which corresponds to the hydrogen-bonded OH vibration of water, is also visible.

### Calibration

For chemometric evaluation the samples were divided into a calibration and a validation set. To enable comparison of the spectral techniques the same sets of samples were used. The calibration set consisted of 96 samples and the validation set of 36 samples. The division into sets was done to obtain similar mean values and standard deviations so that both sets spanned the full range of oil and water content (Table 1).

Several pretreatments of the spectra were investigated for optimization of the calibration model. Mean centering was used in all models, because this technique removes the common information from the spectra. Another possi-

bility of pretreatment is variance scaling (VS) in which each variable is scaled to unit variance. Multiplicative signal correction (MSC) uses a mathematical function to compensate for variations in light scattering. First and second derivatives were calculated because they are useful for extracting band-shift and band-shape features and in eliminating baseline effects. The influence of smoothing using Savitzky–Golay (SG), five-point third-degree polynomial fit, and Norris-derivative (Nd) filter, segment length 5 and a gap between segments 2, was also tested.

The number of factors (latent variables) to be used to construct the models was chosen by cross-validation (employing cancellation groups of four samples) by plotting the number of factors against the root mean square error of cross validation (RMSECV) and determining the minimum. The so obtained RMSECV, which can be used to obtain an estimate of the magnitude of prediction errors, was, furthermore, compared with the root mean square error of calibration (RMSEC). This comparison helps detection of overfitting of the calibration model, which occurs when too many factors are included in the model. This increases the chance that the noise from the calibration data is drawn into the calibration model, which consequently reduces the predictive ability. Inspection of the loading spectra of the factors can also help to exclude factors that are dominated by noise.

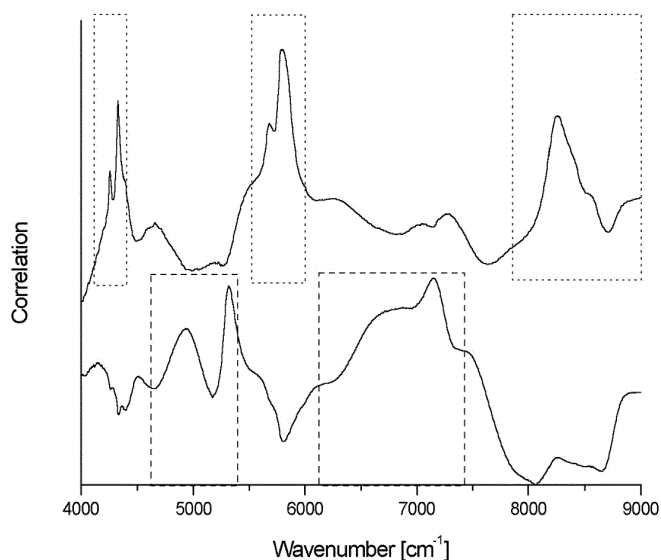
Thorough outlier diagnostics were used to ensure the quality of the calibration sets by plotting the studentized residuals against the leverage of the samples. High leverage indicates that a sample has an abnormally large influence on the model because it has spectral features inconsistent with most of the spectra. If a high leverage sample contains errors or other undesired abnormality this sample may be dangerous and reduce the predictive ability of the resulting calibration model. Such samples were not detected. Especially in Raman measurements samples with high leverage but low errors have been found, and can be attributed to fluorescence, but cannot be considered outliers. They just show another source of variation that must be included in the calibration even if this leads to a more complex model. If not, the predictive ability of the model would be seriously damaged. Samples with high error and low leverage could indicate an error in reference measurement. Such a sample would give bad results in both techniques, Raman and NIR. But this did not occur. We therefore decided not to consider as outlier any sample with high error in only one of the techniques.

Summarizing, the relative performance of a preprocessed data set was assessed by the required number of factors, the RMSEC, and the RMSECV. The predictive ability of the calibration models was evaluated from the root mean square error of prediction (RMSEP) of the independent validation

**Table 1** Characteristics of calibration and validation sets for oil and water content

	Samples	Water content (%)	Mean	stdev <sup>a</sup>	Oil content (%)	Mean	stdev
Calibration set	96	48.3–67.0	59.3	3.7	1.74–3.93	2.62	0.48
Validation set	36	50.0–64.5	59.4	3.9	1.77–3.74	2.52	0.41

<sup>a</sup>Standard deviation



**Fig. 3** NIR Correlation spectra of oil (*above*) and water (*below*). *Dotted lines* mark the regions used for oil calibration and *dashed lines* the regions for water calibration

set. RMSEC, RMSECV, and RMSEP are expressed as weight percentage.

#### Near infrared

One of the advantages of PLS regression is that it is a flexible full spectra method. However, the performance of PLS calibrations can usually be improved by neglecting regions which are unnecessary, because they contain much noise or irrelevant information. The important spectral bands of the analytes (oil and water) can be identified in Fig. 1. Inspection of the correlation spectrum for each of

the analytes (Fig. 3) can also aid to exclusion of unnecessary regions. The correlation spectrum was obtained simply by calculating the correlation of the intensity at every wavenumber in the untreated calibration spectra with the concentrations of the analytes. By combining the spectral information of the analytes and this mathematical approach it was possible to determine three important regions for the determination of oil content, 8890–7815  $\text{cm}^{-1}$ , 6000–5450  $\text{cm}^{-1}$ , and 4410–4175  $\text{cm}^{-1}$ , and two for the determination of water content, 7460–6125  $\text{cm}^{-1}$  and 5375–4550  $\text{cm}^{-1}$ . For both oil and water content the possibility of using the full spectra for calibration was tested but resulted in no better models than that using the chosen regions.

The results obtained for oil and water content calibration in the above given regions are summarized in Table 2. For oil content, already unprocessed spectra gave good results, with an RMSECV of 0.21 and an RMSEP of 0.19. The performance of the calibration models with different pretreatments was very similar. For smoothing the spectra, the very slight reduction in the RMSEC corresponded to selection of eight factors rather than seven. The same tendency was seen in applying MSC. VS, on the other hand, gave the same results as the raw data but with a lower number of factors, six. First-derivative preprocessing led to significant minimization of the number of factors and therefore simplified the models. VS applied to the first-derivative spectra led to an improvement in RMSEC (0.14) but not in predictive ability, probably because of the use of four factors instead of three.

For water content, SG smoothing and VS on the original spectra improved the RMSEP only slightly, from 2.1 to 2.0. Applying MSC led to a model with surprisingly good performance. For this model, ten factors were selected as optimum giving an RMSEC of 1.4, an RMSECV of 2.2, and an RMSEP of 1.9 (results not shown in Table 2). Nevertheless, including three more factors than for un-

**Table 2** Calibration and validation results of NIR models using different spectral pretreatments. The results of models with the best performance are marked in bold

	SG	MSC	VS	1st derivative				2nd derivative		
				SG	Nd	MSC	VS	SG	Nd	
Oil content 4175–4410+5450–6000+7815–8890 $\text{cm}^{-1}$										
Factors	7	8	8	6	<b>3</b>	<b>3</b>	<b>3</b>	4	1	3
R	0.92	0.93	0.93	0.92	<b>0.92</b>	<b>0.90</b>	<b>0.91</b>	0.96	0.87	0.92
RMSEC	0.19	0.18	0.18	0.19	<b>0.19</b>	<b>0.20</b>	<b>0.20</b>	0.14	0.24	0.19
RMSECV	0.21	0.21	0.20	0.21	<b>0.22</b>	<b>0.21</b>	<b>0.21</b>	0.23	0.25	0.22
RMSEP	0.19	0.19	0.20	0.19	<b>0.21</b>	<b>0.21</b>	<b>0.20</b>	0.20	0.24	0.19
Water content 4550–5375+6125–7460 $\text{cm}^{-1}$										
Factors	7	7	7	7	6	6	5	4	4	4
R	0.84	0.84	0.87	0.84	0.95	<b>0.88</b>	0.97	0.94	0.95	0.86
RMSEC	2.2	2.2	2.0	2.2	1.3	<b>1.9</b>	1.0	1.3	1.2	2.0
RMSECV	2.4	2.4	2.6	2.4	2.6	<b>2.4</b>	2.6	2.6	3.5	2.6
RMSEP	2.1	2.0	2.4	2.0	2.1	<b>1.8</b>	2.2	2.0	2.9	2.0

SG=Savitzky–Golay, Nd=Norris derivative, MSC=multiplicative signal correction, VS=variance scaling; R=correlation coefficient, RMSEC=root mean square error of calibration, RMSECV=root

mean square error of cross validation, RMSEP=root mean square error of prediction  
RMSEC, RMSECV, and RMSEP are expressed as weight percentage

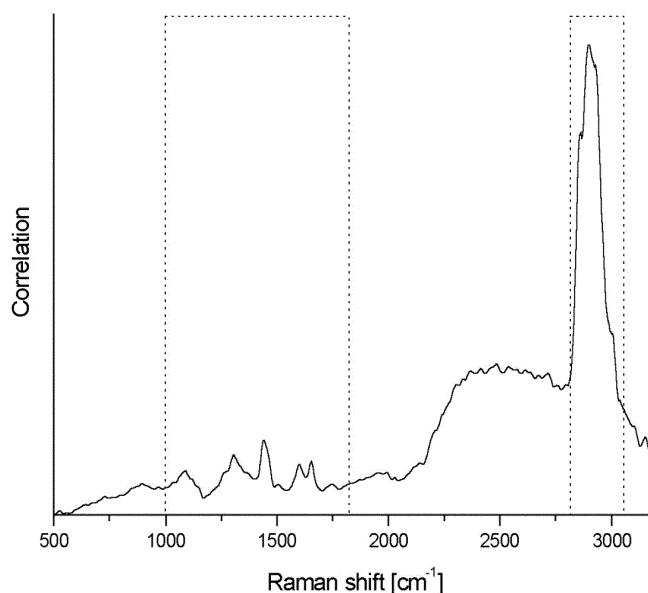
processed spectra to build the model does not seem to make sense, because this kind of correction should minimize scattering effects and not include new variations in the data set. Closer inspection of the factor loadings showed that seven factors already explained more than 99.95% of the spectral variance and 99.9% of the concentration variance. This was, furthermore, confirmed by looking at the loading spectra. The eighth, ninth, and tenth had a very high noise level and a sharp peak at  $7186\text{ cm}^{-1}$ , which took alternating positive and negative sign. Building the same model on seven factors gave an RMSEC of 2.0, an RMSECV of 2.6, and an RMSEP of 2.4, results that can be considered more reliable. First-derivative preprocessing led to reduction of the number of factors, but not as significant as for oil content. Models built with the first derivative gave lower RMSEC values, but the predictive ability did not improve significantly.

The second-order derivative using the SG smoothing filter gave a model for oil content with only one factor. It seems that with this pretreatment the spectral variance was diminished; this, in addition to a much higher noise level, resulted in a bad model. The complexity of the calibration set can be explained by only one factor, thus bad results in terms of fitting and prediction ability were obtained. A similar deficit can be seen in the determination of water, but other than in oil it seems that correlation was found with wavenumbers dominated by noise, thus a minimum in RMSECV was found with four factors. This results in a clear overfitting of this model, which is expressed as a low RMSEC value but very high RMSECV and RMSEP values. In contrast, models built on second-order derivatives smoothed with the Nd filter did not show this problem. For oil content this model yielded similar results, in terms of predictive ability, as the models built on original and first-derivative spectra. For water content, fitting and predictive ability were slightly worse.

The overall performance of oil-calibration models was not significantly affected by the different pretreatments. Models using the first-derivative used a smaller number of factors. They are therefore simpler and can be considered better than those using spectra. Among these the three using SG, Nd, and MSC showed very similar performance; they can, therefore, all be regarded as acceptable. For water content the best results were obtained by use of first derivatives and the Norris derivative-smoothing filter. This model had an RMSECV of 2.4 and an RMSEP of 1.8.

### Raman

Again, before any data pretreatment the spectra (Fig. 2) were inspected as described in the NIR section, to identify unnecessary regions. The correlation spectrum for oil content is presented in Fig. 4. For water content, apart from the region around  $3300\text{ cm}^{-1}$ , the correlation spectra did not have characteristic spectral features, but a strong influence of the fluorescence background. The regions between  $1800$  and  $2600\text{ cm}^{-1}$  and below  $760\text{ cm}^{-1}$  were neglected for both analytes, because they do not contain rel-



**Fig. 4** Raman correlation spectra of oil. Dotted lines mark the regions used for oil calibration

evant spectral information. The regions which contain most information in the oil spectra,  $1000$ – $1800\text{ cm}^{-1}$  and  $2800$ – $3035\text{ cm}^{-1}$ , were chosen for calibration. The calibration models for water content were built on the spectral regions  $760$ – $1800\text{ cm}^{-1}$  and  $2600$ – $3585\text{ cm}^{-1}$ .

The results obtained for oil and water content calibration in the above given regions using different spectral pretreatments are summarized in Table 3. Smoothing had no effect on the error of the model, which suggests that chemometric modeling is not limited by noise. The models with original and smoothed spectra for determination of oil content were built using ten factors. These models had low RMSEC values but high RMSECV and RMSEP, which can indicate overfitting of the calibration. VS worsened the model for both oil and water content. The use of MSC led to models with very limited predictive ability, showing high RMSECV and RMSEP. MSC is a general technique for separating multiplicative (scattering) variations from the additive (chemical) information. The fluorescence background in the Raman spectra, which does not have a multiplicative origin and strongly differs between samples, is a possible reason for the poor performance of the calibration models built on MSC-corrected data. As occurred for near infrared models, first derivative preprocessing led to significant minimization of the number of factors improving, in this instance, the predictive ability of the models. The performance of the smoothing filters was similar. As happened in NIR models, VS applied to the first-derivative spectra improved the fitting but worsened the predictive ability, as can be seen from the RMSECV and RMSEP values. The models built on second-derivative spectra had problems similar to those in NIR calibration, and therefore did not yield better results.

Best results were obtained for models using first derivatives for both oil and water content, probably because the

**Table 3** Calibration and validation results of Raman models using different spectral pretreatments. The results of models with the best performance are marked in bold

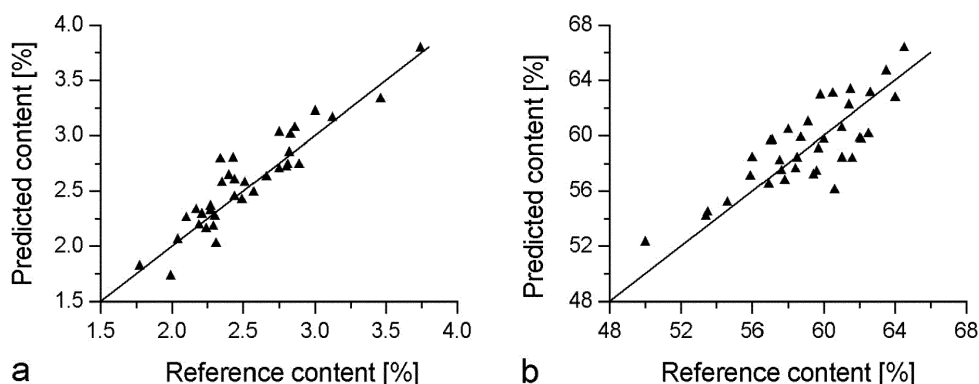
	SG	MSC	VS	1st derivative				2nd derivative		
				SG	Nd	MSC	VS	SG	Nd	
Oil content 1000–1800+2800–3035 cm <sup>-1</sup>										
Factors	10	10	10	8	<b>5</b>	<b>7</b>	4	5	2	6
<i>R</i>	0.94	0.94	0.93	0.90	<b>0.93</b>	<b>0.93</b>	0.84	0.95	0.82	0.92
RMSEC	0.17	0.17	0.18	0.21	<b>0.18</b>	<b>0.18</b>	0.26	0.16	0.28	0.19
RMSECV	0.25	0.25	0.30	0.27	<b>0.27</b>	<b>0.25</b>	0.35	0.28	0.31	0.27
RMSEP	0.24	0.25	0.36	0.24	<b>0.19</b>	<b>0.19</b>	0.29	0.23	0.26	0.23
Water content 760–1800+2600–3585 cm <sup>-1</sup>										
Factors	8	8	8	6	5	<b>5</b>	4	4	2	4
<i>R</i>	0.82	0.81	0.79	0.71	0.90	<b>0.85</b>	0.82	0.91	0.77	0.79
RMSEC	2.3	2.4	2.5	2.8	1.8	<b>2.1</b>	2.3	1.7	2.6	2.5
RMSECV	3.0	3.0	3.0	3.2	3.2	<b>3.0</b>	3.5	3.4	3.4	3.2
RMSEP	2.3	2.3	3.0	2.7	2.1	<b>2.0</b>	2.5	2.3	2.3	2.2

SG=Savitzky–Golay, Nd=Norris derivative, MSC=multiplicative signal correction, VS=variance scaling; *R*=correlation coefficient, RMSEC=root mean square error of calibration, RMSECV=root

mean square error of cross validation, RMSEP=root mean square error of prediction

RMSEC, RMSECV, and RMSEP are expressed as weight percentage

**Fig. 5** Reference vs. predicted concentration of the 36 samples of the independent validation set. Oil content (a) and water content (b) are predicted by PLS models built on the joined Raman and NIR spectra



use of the first derivative eliminated some of the interfering effect of the fluorescence background. These models performed better on the independent validation set, leading to a lower RMSEP – 0.19 and 2.0–2.1 for oil and water, respectively. The calibration models using the Nd smoothing filter were more stable, reflected in the lowest values for RMSECV, 0.25 and 3.0 for oil and water, respectively.

#### Comparison of Raman and NIR model performance

According to the results presented the predictive ability for oil and water content of olive pomace is similar for both techniques. Prediction errors were approximately 0.19–0.20 for oil content and 1.8–2.0 for water content.

The different pretreatments influence the NIR models less than the Raman models. The use of MSC correction, in particular, always led to bad results for Raman models, with high RMSECV and RMSEP values. This can be attributed to the inability of this preprocessing technique to deal with the fluorescence background present in Raman spectra of olive pomace. For both tech-

niques models constructed using first derivatives yielded the best results.

There is a general tendency of Raman models to give higher RMSECV values than NIR models, even if their performance in external validation is similar. This could be because of the higher spectral variation in the Raman data. Characteristics of the samples which are independent of analyte concentration, for example the olive kernel and fluorescent compounds, are much more reflected in the Raman than in the NIR spectra. So, the weight of some samples with very different spectral characteristics might be higher in the model. This can cause a problem in cross validation, when these samples are predicted.

As a general observation, the residual errors for a particular sample are not well correlated between Raman and NIR predictions, which indicates that the Raman and NIR models have substantially different error characteristics. This observation provided the basis for combining the spectral information from both techniques to investigate if better model performance could be achieved by taking advantage of the complementary information of the two techniques.

## Calibration with combined Raman and NIR spectra

The combined NIR and Raman spectra of each sample were used for PLS modeling. The models were built on the same spectral regions, which were identified in the former calibration processes. Only for oil content were the NIR regions 6000–5450  $\text{cm}^{-1}$  and 4410–4175  $\text{cm}^{-1}$  disregarded, because their intensity was very high compared with Raman data. Thus, only the region 7815–8890  $\text{cm}^{-1}$ , which had high correlation with oil content, was used. Again different spectra pretreatments were tested. The best results for oil content were achieved with variance scaling applied to the original spectra. This technique estimates the standard deviation of the intensity values at each wavenumber in the spectra and divides each data point in each spectrum by its estimated standard deviation. This is a simple way of giving every variable an equal chance to contribute to the modeling and in this instance was useful for compensating for the differences in intensity between Raman and NIR signals. The model was built on eight factors and had an RMSEC of 0.18, an RMSECV of 0.21, and an RMSEP of 0.17. This model was more stable, reflected in the lower RMSECV than the Raman model and had a slightly better predictive ability than both Raman and NIR models. For humidity there was no improvement compared with the performance of the models built on the Raman or NIR spectra. The best model was built with the untreated spectra using six factors and had an RMSEC of 2.0, an RMSECV of 2.6, and an RMSEP of 1.9. Figure 5 shows the oil and water content for the independent validation set obtained using the reference analysis plotted against the concentrations predicted by the PLS model based on the combined Raman and NIR spectra.

## Conclusions

This study compared Raman and NIR spectrometry for determination of the oil and water content in olive pomace using the same set of samples. Either of the spectral methods is easier to perform and more expedient than the Soxhlet or NMR methods. According to the results pre-

sented it seems that both techniques should be able to detect excessive levels of oil in olive pomace, arising as a result of anomalies in the extraction process, and to estimate the water content. Both techniques could be implemented in-line in the olive production plant as really non-invasive monitoring techniques, avoiding problems associated with sample handling and pretreatment.

**Acknowledgements** The authors thank C.M. Europa S.A. laboratory for providing the samples and for reference analysis. This study was supported by the University of Jaén under project OB100213. B.M. thanks the Spanish Ministerio de Asuntos Exteriores for a PhD fellowship.

## References

1. Boskou D (1996) Olive oil: chemistry and technology. AOCS Press, Champaign, IL, USA
2. Hassell DC, Bow EM (1998) *Appl Spectrosc* 52:18–29
3. AENOR (1973) UNE 55032 Normas UNE, Asociación Española de Normalización y Certificación (AENOR)
4. Laporte MF, Paquin P (1999) *J Agric Food Chem* 47:2600–2605
5. Tsenkova R, Atanassova S, Toyoda K, Ozaki Y, Itoh K, Fearn T (1999) *J Dairy Sci* 82:2344–2351
6. Tøgersen G, Isaksson T, Nilsen BN, Bakker EA, Hildrum KI (1999) *Meat Sci* 51:97–102
7. Hermida M, Gonzalez JM, Sánchez M, Rodriguez-Otero JL (2001) *Int Dairy J* 11:93–98
8. Adamopoulos KG, Goula AM, Petropakis HJ (2001) *J Food Compos Anal* 14:431–440
9. Archibald DD, Kays SE, Himmelsbach DS, Barton II FE (1998) *Appl Spectrosc* 52:22–31
10. Barthus RC, Poppi RJ (2001) *Vib Spectrosc* 26:99–105
11. Muik B, Lendl B, Molina-Díaz A, Ayora-Cañada MJ (2003) *Appl Spectrosc* 57:233–237
12. Muik B, Lendl B, Molina-Díaz A, Ayora-Cañada MJ (2003) *Anal Chim Acta* 487:211–220
13. García Mesa JA, Hermoso Fernández M, Cáceres Alonso P (1996) *Grasas y aceites* 47:317–322
14. AENOR (1973) UNE 55031 Normas UNE, Asociación Española de Normalización y Certificación (AENOR)
15. Hourant P, Baeten V, Morales MT, Meurens M, Aparicio R (2000) *Appl Spectrosc* 54:1168–1174
16. Sadeghi-Jorabchi H, Hendra PJ, Wilson RH, Belton PS (1990) *JAOCS* 67:483–486