Oil accumulation in intact olive fruits measured by near infrared spectroscopy – acousto-optically tunable filter.

Running title: Oil content in olive fruits using NIRS

Andrea Bellincontro [1*], Giovanni Caruso [2], Fabio Mencarelli [1], Riccardo Gucci [2]

[1]Dept. Innovazione dei Sistemi Biologici, Agro-alimentari e Forestali, University of Tuscia, Via San Camillo de Lellis snc, 01100 Viterbo, Italy. phone 0039 0761 357313, 0039fax 357498

[2]Dept. Coltivazione e Difesa delle Specie Legnose, University of Pisa, Via del Borghetto 80, 56124 Pisa, Italy

*Corresponding author: Andrea Bellincontro, [1]Dept. for Innovation in Biological Agro-food and Forest systems (DIBAF) –Postharvest Laboratory, University of Tuscia.

e-mail: bellin@unitus.it

Abstract

BACKGROUND: A field experiment was conducted to test the reliability of the near infrared spectroscopy (NIR) method to measure mesocarp oil content *in vivo* against nuclear magnetic resonance (NMR) determinations using three different olive cultivars at different stages of ripening.

RESULTS: In PLS model carried out for the cultivar Arbequina the R^2c (coefficients of determination in calibration) obtained was of 0.991, while the R^2cv (coefficients of determination in cross-validation) of 0.979; for the cultivar Frantoio the indexes were respectively of 0.982 and 0.971, for the cultivar Leccino of 0.977 for R^2c and 0.965 for R^2cv. Finally, for the combined model (sum of the three varieties) those indexes were respectively equal to 0.921 and 0.903. The RPD (Residual Predictive Deviation) ratio was insufficient for predictive model of cultivar Leccino only (1.98), whereas in other cases RPD ratios were completely sufficient, within the estimation range over 2.5 - 3 (2.61 in global model, and 4.23 in cultivar Frantoio), or describing a great capacity with values greater than 5 as in the case of the cultivar Arbequina (9.58).

CONCLUSION: The NIR proved a novel, rapid, reliable method to monitor the oil accumulation process in intact olive fruits in the field. The innovative approach of coupling NIR and NMR technologies opens new scenarios for determining the optimal time for harvesting olive trees to obtain maximum oil production.

Keywords: *Olea europaea* L., cultivar, mesocarp, NIRS, NMR, PLSR.

INTRODUCTION

The olive fruit is a drupe, composed of an external exocarp, a fleshy mesocarp, and a woody endocarp surrounding the seed. The mesocarp is the largest tissue as it accounts for 60-70% of the dry weight (DW) of entire fruits respectively, depending on cultivar and water availability[1,2] The mesocarp is the tissue of economic value not only for table consumption but also for oil production since up to 98% of the oil accumulates there. About 90% of olive fruits, produced over 10 million ha worldwide, are destined to oil production, while only 10% is for table consumption.[3]

The process of oil accumulation in olive fruits starts appreciably in the second half of the summer, progresses rapidly for about eight weeks and then slows down as the fruit approaches ripening.[1,4] The cultivar and climatic conditions during fruit development strongly influence oil accumulation and determine the oil content at harvest.[4] During ripening the olive fruit undergoes modifications in texture, colour, and chemical composition[5], that are also the result of interactions between environmental and genetic factors. For example, soil water availability alters the progression of ripening and eventually affects oil accumulation and quality.[6,7] Thus, in order to determine the optimal stage at which olive fruits should be harvested, several parameters are to be monitored.

Determining the oil content and composition using analytical methods is time consuming and expensive and, for this reason, simple methods have been proposed to guide olive growers in the decision making process about harvesting. The colour assessment of the exocarp and the mesocarp is a quick method often used in the field[8], but the correlation between the degree of pigmentation of both tissues and either oil accumulation or quality parameters is cultivar dependent[9] and varies with prevailing

environmental conditions during fruit development. Colour change occurs because of chlorophyll degradation and accumulation of anthocyanins that turn the skin to more or less dark brown hues depending on anthocyanins concentration.[5,10] Temperature and light play a key role in biosynthesis and accumulation of anthocyanins, phenolic compounds and flavonoids in grape berries, although individual effects of these environmental factors are difficult to separate.[11] Thus, it is not surprising that the relationship between colour, oil content, and composition in olive fruits is often misleading and, therefore, determining harvest time for maximum oil quality and oil content based on fruit colour is unsatisfactory according to modern standards of olive growing.

Visible and near infrared spectroscopy (Vis-NIRS) is a well known technique for the non-destructive measurement of quality attributes of food commodities including fresh fruits and vegetables.[12] The NIR region contains information concerning the relative proportions of C-H, N-H, and O-H bonds, which are the primary structural components of organic molecules, and so NIRS can be used, combined with a chemometric approach, as an alternative technique to the destructive, analytical methods commonly used for quality assessment. Partial least squares (PLS) regression is the chemometric application[13,14] usually employed in the development of predicting models of qualitative attributes for fruits and vegetables.[15]

Regarding the use of NIRS on olive fruits and related products a few studies were recently performed[16-20]. Cayuela et al.[21] (2009) used an application of the NIR-acousto-optic tunable filter (AOTF) for the prediction of fruit moisture, free acidity, and oil content in intact olive fruits. The potentiality of the same device in predicting total and specific olive phenols through the ripening evolution was explored by Bellincontro

et al.[22] Moreover, the NIR-AOTF apparatus was tested for the prediction of the percentage of oil content in two olive cultivars and two selections.[23] The high performance of the NIR-AOTF equipment is mainly due to the wavelength selection capability of the acousto-optical filter.[24]

The objective of the present study was to test and validate the potential use of NIR-AOTF for the determination of the oil content of olive fruits during their development. For this reason, we calibrated the NIR-AOTF against the NMR technique, the latter being the most rapid and reliable method to measure the mesocarp oil content in fruits samples. In order to have a wide range of oil content values the study was conducted on three olive cultivars with different patterns of oil accumulation at different stages of fruit development.

MATERIALS AND METHODS

Plant material and olive fruit sampling

An irrigated olive orchard growing in a sandy-loam soil at the experimental farm of University of Pisa (43° 01’ N; 10° 36’ E) was used in 2010. Fruits were sampled from four fully-productive trees per cultivar of ‘Frantoio’, ‘Leccino’ and ‘Arbequina’, which yielded an average of 24, 18 and 18 kg per tree, respectively. All trees from that orchard were harvested on 25 October (164 days after full bloom), except those that were used for sampling fruits at later dates of the current study. The three cultivars are widely cultivated worldwide because of their agronomic or qualitative characteristics. ‘Frantoio’ is renowned for the excellent quality of the oil, ‘Leccino’ for high yields and its adaptation to cold climate conditions, ‘Arbequina’ for its suitability to very high density plantations. Fruits of cultivars ‘Pendolino’ and ‘Moraiolo’ from that same

orchard were also used to either calibrate the response of NMR equipment or determine the minimum duration of the drying period prior to NMR analysis, as explained in the next paragraph.

The orchard floor was permanently covered with grass and water was supplied by subsurface drip irrigation.[25] An irrigation experiment, consisting of three levels (full irrigation, 50% deficit irrigation, complementary irrigation) was established since 2006 and maintained until 2010. In 2010 water was distributed from July 9 through September 17 to satisfy fully (1997 m^3 ha^{-1}) or partially (175 and 92 m^3 ha^{-1}) tree water needs. Relatively low volumes of water were needed for the deficit treatments because of frequent precipitations that occurred during the summer. Annual precipitation in 2010 was 1185 mm, 25% of which from July 1 through October 25, much higher than the 18-year mean (1990-2008) of 635 mm for that site.

Intact, healthy fruits were sampled for the determination of oil content by NIR-AOTF and NMR. Twenty fruits were randomly sampled from around the canopy of each tree at six (‘Frantoio’ and ‘Leccino’) and four (‘Arbequina’) dates from the beginning of October through the end of November. Totally, twenty five sample set of drupes were collected for Leccino cultivar, while for Frantoio and Arbequina cultivars were collected respectively twenty and sixteen set of samples. Each sample set was represented by 20 olives, and the same fruits were firstly spectral detected and then destined to the analytical measurements. Prior to the determination of oil content the maturation index (MI) was measured according to a standard methodology, whereby the skin and flesh colours were scored according to a 0 to 7 scale.[8]

NMR determination of mesocarp oil content

The mesocarp oil content was measured using an NMR Oxford MQC-23 analyzer (Oxford Analytical Instruments Ltd., Oxford, UK) at 40 °C. The NMR equipment allows to measure the oil content in small samples of fruit dry tissue (about 4 g) rapidly (16 sec for each reading) and, hence, to process a large number of samples in a short period of time.[26] In general, we followed the method by Del Rio and Romero[26] modified as follows. Samples of 20 fruits were taken to the laboratory, the mesocarp separated from the endocarp and dried in an oven at 70 °C for 48 h. A preliminary time-course experiment conducted on mesocarp tissue and milled fruits of three cultivars ('Frantoio', 'Leccino' and 'Moraiolo'), taken from the same orchard, had shown that 48 h was a sufficient time to dry the samples adequately and to obtain stable values of DW (dry weight) and oil content (**Figure 1**). Successively, 4-5 g of dry pulp were cut in small (2 to 5 mm) pieces by a razor blade, put in a 5-mL glass vial and kept at 70 °C until all samples had been prepared. Prior to analyses, samples were conditioned at room temperature for 15 min and then their oil content measured in triplicate.

The NMR MQC-23 analyzer was calibrated against known standards and olive fruit samples, the oil content of which had been determined by the Soxhlet method[27] using 1000 g of fruits from a single tree of three cultivars ('Frantoio', 'Leccino' and 'Pendolino') growing in the same orchard and harvested at different dates in 2009. The fruits were washed with tap water, their surface dried with blotting paper, then crushed using a hammer mill (MM-100, MC2, Ingenieria y Systemas, Sevilla, Spain) and the paste oven-dried at 70 °C for 72 h. An aliquot (4-5 g) from each subsample was used to determine the oil content of the milled olive fruits.

Spectral acquisition, chemometric procedure and data analysis

A Luminar 5030 miniature, hand-held NIR analyzer (Brimrose Corporation, Baltimore, Maryland, USA), based on the AOTF-NIR principle, was used for spectral detection. This portable device can be directly used in the field, but in the current study spectral detections were conducted under laboratory conditions. Two different measurements were performed on each intact olive fruit through contact between the external gun of the NIR device and the epicarp in correspondence to the equatorial zone of the drupe using the diffuse reflectance method of detection, while the raw spectra were detected and recorded in transmittance mode.[22] A single measurement, at the speed of 16,000 wavelength sec^{-1}, was conducted in the 1100–2300 nm range, with 2 nm wavelength increments and 10 spectra per average, wich represents a good compromise between speed of acquisition and signal quality of the spectrum. The average of the two measurements, carried out on the opposite faces of the drupe, was the spectral response of the fruit. A total number of 320, 460, and 500 drupes were spectrally tested for 'Arbequina', 'Frantoio', and 'Leccino', respectively.

Raw spectra were statistically pre-treated for absorbance (log 1/T) transformation using SNAP! 2.03 software (Brimrose). Before the calibration and construction of the prediction models, the spectral variations of the data sets were analyzed through Principal Component Analysis (PCA). The absorbance spectra, obtained as spectral average of each olive subsets, were used as X-variables for the final models. Mean normalization, Multiplicative Scattering Correction (MSC), and Standard Normal Variate (SNV) treatments, first order of Savitzky-Golay filter (6 points of smoothing) or second order of Savitzky-Golay filter (6 points of smoothing) were also tested, although they were not utilized in final modeling. In fact, absorbance spectra, without any data pre-treatment, were identified as most effective in achieving the goal

of the model calibrations. Partial Least Squares (PLS) models[28] were obtained on the full spectrum (1100-2300 nm), considering the spectral significant variables at specific wavelength intervals. The mean values ± standard deviation (SD) obtained by the reference measurements were used as Y-variables in the PLS matrices in which they were opposed to the averaged spectra, as reported below. Predicting models of the percentage of oil were developed for individual cultivars, and then a global model was calculated starting from the data sum of the three sets and also included in predicting models. The total sample set of data was considered for the calibration and validation procedure, which was carried out by leave-one-out cross-validation method. Outliers identification and elimination was not performed. The statistical indexes R^2_C (coefficient of multiple determination of calibration) and R^2_{CV} (coefficient of multiple determination of cross-validation), Standard Error of Calibration (SEC), Root Mean Standard Error of Calibration (RMSEC), Root Mean Standard Error of Cross Validation (RMSECV) and Bias were used to determine the significance of the calculations. The RPD (Ratio of Performance to Deviation） ratios, defined as the ratio between the SD and the Standard Error of Cross-Validation (SECV)[29], were also calculated to derive the final models for the all parameter estimation. Statistical pre-treatments, PCA, and PLS models were performed using Unscrambler v9.7 software (CAMO ASA, Oslo, Norway); graphs, score plot and scatter plots were performed, after data exportation from Unscambler, using SigmaPlot v. 11.0 (Systat Software Inc., San Jose, CA, USA).

RESULTS AND DISCUSSION

Fruits from the three cultivars had different MI and FW during the sampling period. Fruits of ‘Frantoio’ and ‘Arbequina’ were green until the end of October,

whereas by that time those of ‘Leccino’ had already reached a score of 3.5, corresponding to dark colour over about 50% of the epicarp (**Figure 2**). By the end of October the skin of ‘Leccino’ fruits was completely black, that of ‘Frantoio’ was only 50% black. ‘Leccino’ and ‘Frantoio’ fruits had similar FW, those of ‘Arbequina’ were lighter. The range of oil concentrations was wide during fruit maturation and across cultivars. The mesocarp of ’Frantoio’ fruits had higher oil content than ‘Leccino’ or ‘Arbequina’ fruits throughout the sampling period (**Figure 2**). Besides genetic factors, cultural practices (e.g. irrigation) or the cropping condition of the tree can affect mesocarp oil content.[6,30] For instance, the oil content of cvs. Leccino and Morisca increased as the degree of water deficit experienced during fruit development decreased.[6,30] Moreover, it has been shown that high crop loads determined a decrease in fruit oil content, that was more pronounced as water deficit increased[30], and that the oil content increased by 10-15% when the crop load was halved.[6,31]

The absorbance mean raw spectra and in 2nd derivative, relative to all data acquired for all fruits of individual cultivars are reported respectively in **Figure 3 A and B**. The first band occurred at 1150 nm, and corresponded to a combination of the symmetric and asymmetric OH stretching and OH bending bands. A second band appeared at 1200 nm, corresponding to the second overtone of the CH stretching vibrations of CH_3, CH_2, and CH=CH. Spectra were mainly dominated by two principal water absorption bands of around 1450 nm and 1920-1950 nm.[32] They represented the first overtone of the symmetric and asymmetric OH stretching and combination bands (1450 nm), and the combination of the OH stretching band and to the OH bending band (1920-1950 nm), respectively.[33,34] The two bands at 1720 and 1750 nm corresponded to the first overtone of the CH stretching vibration of CH_3, CH_2 and CH=CH. The last band

in olive fruit spectra, observed at 2250 nm, was due to the combination of the CH stretching vibrations of the CH_3, CH_2 with other vibrations. The peaks at 1200, 1720-1750, and 2250 nm can be attributed to the presence of oil, as reported previously.[35] In PLS modeling for percentage of oil calibrations and predictions, specific wavelengths were not selected and the entire spectrum (1100-2300 nm) was included in chemometric calculations, because of the greater contribute on the modeling. Preliminary PCA, carried out on all the spectral detections, was used just for sample description, while outlier selection was not applied and no samples were discarded. The same PCA reported in **Figure 4** showed that virtually all of the variance was explained by PC1 and PC2 (respectively 74% and 22%), while for the explanation of total residual variance (99%) four PCs were required. In all procedures that require chemometric approaches, a high variability in the concentration of the parameters, measured by destructive measurements, is relevant for successful modeling. In our study mesocarp oil content values ranged over a wide interval and statistical index showed high variability (**Table 1**), which is good for the accuracy and the robustness of the final models that are largely influenced, as well as the effectiveness of the spectral detections, by the variability of the destructive values.

The raw spectra (only transformed in absorbance, log 1/T) were used for modeling since they did not require any particular filtering and were more effective than spectra sets subjected to different pretreatments (data not shown, see also Materials and Methods). Calibration and cross-validation results for all models, in terms of estimated mesocarp oil content with respect to the olive varieties, are reported in Table 2. Scatter plots obtained from the same data sets showed high correlations for all the models of the tested olive cultivars, including total model carried out as the sum of three varieties

257 employed (**Figure 5**). In the PLS model developed for the cultivar Arbequina the R^2c
258 (coefficients of determination in calibration) obtained was 0.991, whereas the R^2cv
259 (coefficients of determination in cross-validation) was 0.979; for the cultivar Frantoio
260 the indexes were respectively 0.982 and 0.971, for the cultivar Leccino 0.977 and 0.965
261 for R^2c and R^2cv, respectively. Finally, for the combined model (sum of the three
262 varieties) those indexes were respectively equal to 0.921 and 0.903.

263 R^2cv values greater than 0.9 are usually considered to provide good quantitative
264 information for the estimation of the predictive accuracy of models.[32] Our R^2c and R^2cv
265 values were also higher than those obtained for the oil content of two Spanish olive
266 cultivars (Arbequina and Picual) in a similar study where the NIR-AOTF equipment
267 was used in combination with NMR that was the reference method.[23] In another study,
268 in which NIR detection was obtained by a different spectral device but the percentage of
269 oil content was measured by the same NMR method, R^2 in calibration and prediction
270 were 0.98 and 0.96, respectively.[35] Therefore, we can consider the R^2 indexes as the
271 parameter able to explain the good correlation between the non-destructive opposed to
272 the destructive data set. The high correlation obtained in this work can be probably
273 attributed to the accuracy and precision of the reference data of mesocarp oil content, as
274 well as to the effectiveness of the NIR-AOTF. Cozzolino et al.[36] emphasized the
275 importance of measuring accurately the destructive samples in chemometric studies.
276 However, the real and applicative performance of the predictive models is better defined
277 if combined with the estimation indexes of the potential errors in calibration and
278 prediction or cross-validation (RMSEC and RMSECV). In our study RMSEC and
279 RMSECV values were respectively 0.276% and 0.426% for cultivar Arbequina, 0.668%
280 and 0.849% for cultivar Frantoio, 0.87% and 1.069% for cultivar Leccino (**Table 2**). As

281 for the global model, the values of the resulting RMSEC and RMSECV reached 1.413%
282 and 1.601%. In their work, Gracia and Léon[23] reported RMSECV ranging between 1.52
283 and 1.89%, while Dupuy et al.[35] obtained standard errors in prediction (SEP) equal to
284 1.18, which was slightly lower (0.78 %) when the NIR spectroscopy was combined with
285 MIR. Dardenne[37] clarified how the cross-validation method can be used in NIRS
286 applications in order to predict qualitative attributes, even if the accuracy of the methods
287 could be improved by the use of an appropriate, preferably external, set of validation.
288 However, in leave-one-out cross validation, one sample is removed from the dataset and
289 a calibration model is built on the base of the remaining subset. Removed samples are
290 then used for the calculation of the prediction residual[36]. This validation method can be
291 considered satisfactory especially when, as in our case, the experimental dataset is
292 limited and it is not possible to arrange it into two separate subsets: the big one to be
293 used for calibration and the small one for external validation. In our case we started
294 from a total sample of 1280 drupes and 2560 spectral measurements to end with a
295 maximum of 192 reference values of oil content (**Table 2**), due to the need to pool 20
296 olive fruits in order to obtain a representative number of samples for the NMR analyses.
297 However, when we tried to use the dataset arranged for the global model by separating
298 it into two subsets (calibration and prediction) for the consequent PLS modeling, the
299 results were not significant as in the case of the cross validated model (data not shown).
300 Certainly, a validation procedure of the predicting models by on-field determination of
301 the oil content will be carried out. The RPD ratio is a statistical index used to evaluate
302 the predictive ability of the NIR applications between low and high values of the
303 response variable.[29,38] Values higher than 5 indicate good discrimination especially if
304 destined to quality and food control.[38,39] In our study the RPD ratios were completely

sufficient, within the estimation range over 2.5 - 3 (2.61 in global model, and 4.23 in cultivar Frantoio), or describing a great capacity with values were greater than 5 as in the case of the cultivar Arbequina (9.58); only in the case of cultivar Leccino an insufficient RPD ratio (1.98) was obtained for predictive model.

In conclusion, traditional techniques to determine oil content and oil composition are time consuming and their responses are too slow to be compatible with the objective of identifying optimal harvest time in the field. NMR is a reliable technique already available to determine the oil content of few grams of entire fruits or mesocarp tissue, but it implies destructive sampling and drying of fruit specimens (Del Rio and Romero, 1999).[26] The robustness and high correlations obtained in this study demonstrate the feasibility and potential of NIR-AOTF spectroscopy, coupled with NMR technique for calibration, as a reliable method to measure the oil content of entire olive fruits. This fast technique proved adequate for monitoring the oil accumulation during fruit ripening and provide useful information on the best time for harvesting olive fruits and can significantly enhance the confidence on developing indexes for best harvesting of olives.

ACKNOWLEDGEMENTS

We thank Michele Bernardini and Maurizio Gentili for excellent technical assistance. The irrigation system was kindly supplied by Netafim Italia. Research supported by UNAPROL-Italy projects Reg. UE no. 2080/2005 and no. 867/2008.

328 **REFERENCES**

329 1. Gucci R, Lodolini EM and Rapoport HF, Water deficit-induced changes in mesocarp
330 cellular processes and the relationship between mesocarp and endocarp during
331 olive fruit development. *Tree Physiol* **29**:1575-1585 (2009).

332 2. Lavee S, Hanoch E, Wodner M, Abramowitch E, The effect of predetermined deficit
333 irrigation on the performance of cv Muhasan olives (Olea europaea L.) in the
334 eastern coastal plain of Israel. *Sci Hortic* **99**:267-277 (2007).

335 3. FAOSTAT 2010. On-Line Statistical Database of the Food and Agricultural
336 Organization of the United Nations. http://apps.fao.org/ [4 October 2011].

337 4. Tombesi A, Olive fruit growth and metabolism. *Acta Hortic* **356**:225-232 (1994).

338 5. Conde C, Delrot S and Geros E, Physiological, biochemical and molecular changes
339 occurring during olive development and ripening. *J Plant Physiol* **165**(15):1545-
340 1562 (2008).

341 6. Gucci R, Lodolini EM and Rapoport HF, Productivity of olive trees with different
342 water status and crop load. *J Hortic Sci Biotechnol* **82**(4):648-656 (2007).

343 7. Servili M, Esposto S, Lodolini E, Selvaggini R, Taticchi A, Urbani S, MOntedoro
344 GF, Serravalle M, Gucci R, Irrigation effects on quality, phenolic composition
345 and selected volatiles of virgin olive oil cv. Leccino, *J Agric Food Chem*
346 **55**:6609-6618 (2007).

347 8. Beltran G, Uceda M, Hermoso M and Frias L, Maduracion, in *El Cultivo del Olivo*,
348 ed. by Barranco D, Fernández-Escobar R, Rallo L (Mundi-Prensa), Madrid, pp.
349 159-183 (2004).

350 9. Servili M, Montedoro GF, Pannelli G, and Famiani F, Influenza delle variabili
351 pedologiche, tecnologiche e varietali sulla qualità degli oli vergini di oliva. *Atti*

352 *convegno "Problematiche qualitative degli oli vergini di oliva"*, Sassari 6
353 novembre 231-245 (1990).
354 10. El Riachy M, Priego-Capote F, Leon L, Rallo L, and Luque de Castro MD,
355 Hydrophilic antioxidants of virgin olive oil. Part 2: biosynthesis and
356 biotransformations of phenolic compounds in virgin olive oil as affected by
357 agronomic and processing factors. *Eur J Lipid Sci Technol* **113**:692-707 (2011)
358 11. Downey MO, Dokoozlian NK, and Krstic MP, Cultural practices and environmental
359 impacts on the flavonoid composition of grapes and wine: a review of recent
360 research. *Am J Enol Vitic* **57**(3):257-268 (2006)
361 12. Lin H, and Ying Y, Theory and application of near infrared spectroscopy in
362 assessment of fruit quality: a review. *Sens & Instrumen Food Qual* **3**:130–141
363 (2009).
364 13. Geladi P, Chemometrics in spectroscopy Part I. Classical chemometrics.
365 *Spectrochimica. Acta Part B*, **58**:767-782 (2003).
366 14. Wold S, Sjöström M, and Eriksson L, PLS-regression: a basic tool of chemometrics.
367 *Chemom Intell Lab Syst* **58**:109–130 (2001).
368 15. Nicolai B, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, and Lammertyn J,
369 Nondestructive measurements of fruit and vegetable quality by means of NIR
370 spectroscopy: a review. *Postharvest Biol Technol* **46**:99-118 (2007).
371 16. Leon L, Garrido A, Downey G, Parent and harvest year effects on near-infrared
372 reflectance spectroscopic analysis of olive (Olea europaea L) fruit traits. *J Agric*
373 *Food Chem* **52**:4957–4962 (2004).

17. Bendini A, Cerretani L, Di Virgilio F, Belloni P, Lercker G and Gallina T, In process monitoring in industrial olive mill by means of FT-NIR. *Eur J Lipid Sci Technol* **109**:498–504 (2007).

18. Barros AS, Nunes A, Martins J and Delgadillo I, Determination of oil and water in olive and olive pomace by NIR and multivariate analysis. *Sensing and Instrumentation for Food Quality and Safety* **3**(3):180-186 (2009).

19. Cayuela JA and Pérez-Camino MC, Prediction of quality of intact olives by near infrared spectroscopy. *Eur J Lipid Sci Technol* **112**:1209-1217 (2010).

20. Dupuy N, Galtier O, Le Dréau Y, Pinatel C, Kister J and Artaud J, Chemometric analysis of combined NIR and MIR spectra to characterize French olives. *Eur J Lipid Sci Technol* **112**:463-475 (2010).

21. Cayuela J A, García J M and Caliani, N, NIR prediction of Fruit moisture, free acidity and oil content in intact olives. Grasas y Aceites **60**:194–202 (2009).

22. Bellincontro A, Taticchi A, Servili M, Esposto S, Farinelli D, Mencarelli F, Feasible application of a portable NIR-AOTF tool for on-field prediction of phenolic compounds during ripening of olives for oil production. *J Agric Food Chem* **60**(10):2665-2673 (2012)

23. Gracia A and Léon L, Non-destructive assessment of olive fruit ripening by portable near infrared spectroscopy. *Grasas y Aceites* **62**(3):268-27 (2011).

24. Barbieri Gonzaga F, and Pasquini C. Near-Infrared emission spectrometry based on an acousto-optical tunable filter. *Anal. Chem*. **77**:1046-1054 (2005).

25. Caruso G, Rapoport HF and Gucci R, Long-term evaluation of yield components of young olive trees during the onset of fruit production under different irrigation regimes. *Irr Sci* DOI: 101007/s00271-011-0286-0 (2011).

398 26. Del Rio C and Romero AM, Whole, unmilled olives can be used to determine their
399 oil content by nuclear magnetic resonance. *HortTechnology* **9**(4):675-680
400 (1999).
401 27. Lynk WE, Official and tentative methods of the American Oil Chemists' American
402 Oil Chem Soc, Champaign, Ill P (1975)
403 28. Wold S, Sjöström M, and Eriksson L. PLS-regression: a basic tool of chemometrics.
404 *Chemom. Intell. Lab. Syst.* **58:**109–130 (2001)
405 29. Williams PC and Sobering DC, How do we do it: A brief summary of the methods
406 we use in developing near infrared calibrations, In *Near infrared spectroscopy:*
407 *The future waves,* ed. by Davies AMC and Williams PC. Chichester: NIR
408 Publications, pp. 185-188 (1996).
409 30. Martin-Vertedor AI, Perez-Rodriguez JM, Prieto, H and Fereres E, Interactive
410 responses to water deficits and crop load in olive (Olea europaea L, cv Morisca)
411 II: Water use, fruit and oil yield. *Agric Water Manage* **98**:950-58 (2011).
412 31. Barone E, Gullo G, Zappia R and Inglese P, Effect of crop load on fruit ripening and
413 olive oil (Olea europaea L). *J Hortic Sci* **69**:67-73 (1994).
414 32. Shenk JS and Westerhaus MO, Calibration the ISI way. In *Near infrared*
415 *spectroscopy: The future waves,* ed. by Davies AMC and Williams PC.
416 Chichester: NIR Publications, pp. 198–202 (1996).
417 33. Maeda H, Ozaki Y, Tanaka M, Hayashi N, Near spectroscopy and chemometric
418 studies of temperature dependent spectral variations of water: Relationship
419 between spectral changes and hydrogen bonds. *J Near Infrared Spec* **3**:191–201
420 (1995).

34. Bertrand D, Spectroscopie de l'Eau, in *La spectroscopie infrarouge et ses applications analytiques*, 2nd Edn, ed. by Tec & Doc, Paris, pp. 94–104 (2006).

35. Dupuy N, Galtier O, Le Dréau Y, Pinatel C, Kister J, Artaud J. Chemometric analysis of combined NIR and MIR spectra to characterize French olives. *Eur. J. Lipid Sci. Technol*. 112:463-475 (2010).

36. Cozzolino D, Cynkar WU, Shah N, and Smith P, Multivariate data analysis applied to spectroscopy: Potential applications to juice and fruit quality. *Food Research International* **44**:1888-1896 (2011).

37. Dardenne P, Some considerations about NIR spectroscopy: Closing speech at NIR-2009. *NIR News* **21**(14):8-9, (2010).

38. Williams PC, Implementation of near-infrared technology, in *Near Infrared Technology in the Agricultural and Food Industries*, ed. by Williams PC and Norris KH. St Paul, USA: American Association of Cereal Chemist, Minnesota pp.145-169 (2001).

39. Fearn T, Assessing calibration: SEP, RPD, RER and R2. *NIR News* **13**:12-14 (2002).

Table 1. Mean, standard deviation (SD), and range of mesocarp oil content (% dry weight) measured by NMR for total sample sets and individual olive cultivars (Arbequina, Frantoio, and Leccino).

	Arbequina	Frantoio	Leccino	Total
Fruits (n)	320	460	500	1280
Data set (n)	48	69	75	192
Mean	59.1	64.7	61.0	61.8
SD	55.25	59.05	52.41	52.41
minimum	62.8	72.0	68.2	72.0
maximum	2.1	3.6	4.1	4.2

Table 2. Calibration and cross validation results relative to the PLS models obtained for three different olive cultivars and data set. Number of data included in the dataset (n), coefficient of determination in calibration (R^2_c) and in cross-validation (R^2_{cv}), root mean standard error in calibration (RMSEC) and in cross-validation (RMSECV), standard error of calibration (SEC), bias, number of latent variables (LVs), and ratio of performance to deviation (RPD) are reported.

Cultivar (n)	Calibration					Cross validation			
	R_c^2	RMSEC	SEC	Bias	LVs	R_{cv}^2	RMSECV	SECV	RPD
Arbequina (48)	0.991	0.276	0.279	-3.18E-06	11	0.979	0.426	0.43	9.58
Frantoio (69)	0.982	0.668	0.673	2.27E-06	8	0.971	0.849	0.855	4.23
Leccino (75)	0.977	0.87	0.876	-6.61E-06	8	0.965	1.069	1.076	1.98
Total (192)	0.921	1.413	1.417	-2.98E-07	11	0.903	1.601	1.605	2.61

Captions for figures

Figure 1. Oil content (A) and weight (B) of mesocarp tissue (open symbols) and milled fruits (closed symbols) of three olive cultivars (‘Frantoio’ ○ - ●, ‘Moraiolo’ □ - ■, ‘Leccino’ Δ - ▲) measured after different drying periods at 70 °C. Fruits were crushed using a hammer mill (MM-100, MC2, Ingenieria y Systemas, Sevilla, Spain).

Figure 2. Maturation index, fruit fresh weight and oil content of fruits sampled from trees of ‘Frantoio’ and ‘Leccino’ from the end of September through the end of November 2010. Fruits of ‘Arbequina’ were sampled only until the end of October. Symbols are means of four replicates - standard error bars.

Figure 3. NIR-AOTF mean spectra, raw (A) and in 2nd derivative (B), of all olive samples measured during their ripening evolution. Spectra are plotted as absorbance units calculated from the original detections (log 1/T) *versus* the wavelength (nm). Significant bands are indicated.

Figure 4. Three-dimensional score plot of the principal component analysis (PC1 vs PC2 vs PC3) carried out on the absorbance NIR-AOTF spectra of grouped samples generated using three olive cultivars (Arbequina, Frantoio, and Leccino). Percentage of the explained variance is reported in brackets on each axis.

Figure 5. Scatter plots relative to the PLS models for percentage of oil prediction carried out on cultivar Arbequina (A), Frantoio (B), Leccino (C), and on the global dataset of olive samples (sum of the three cultivar) (D). For each cultivar measured values are plotted *versus* predicted values and calibration and validation datasets are reported.

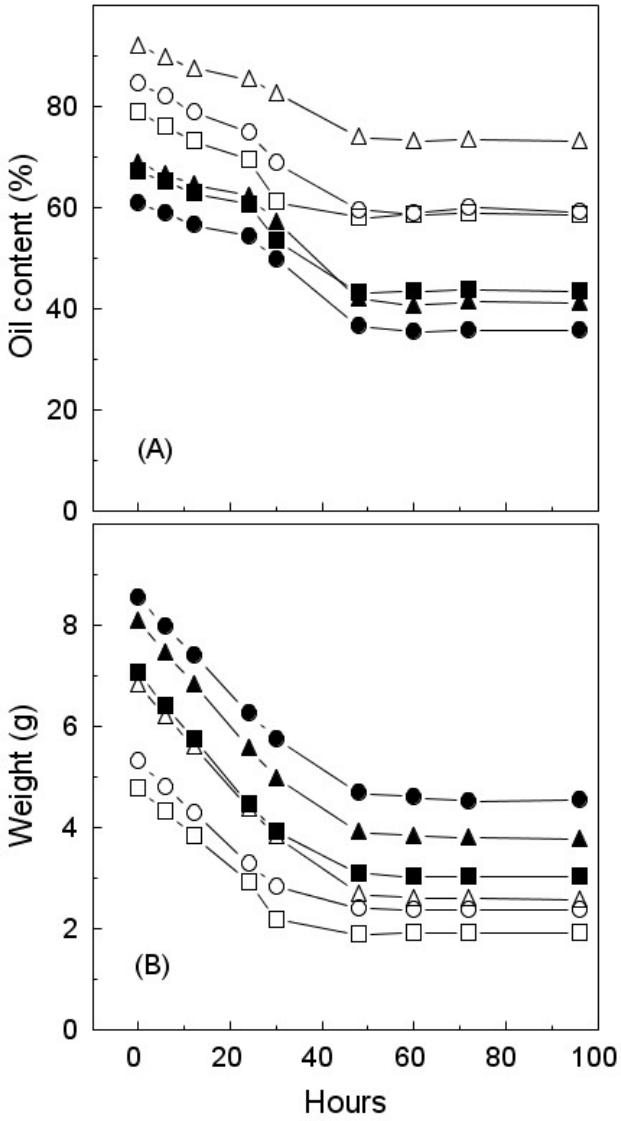
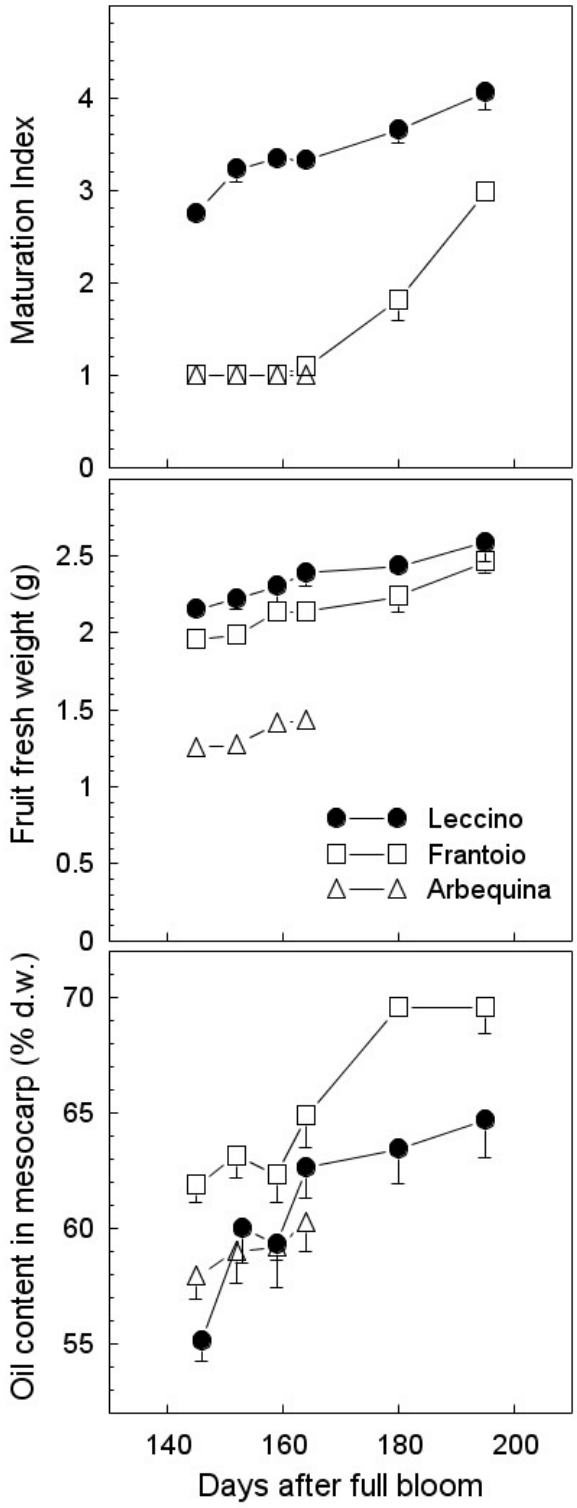

Oil content (%)
0
20
40
60
80
(A)
Weight (g)
0
2
4
6
8
(B)
0
20
40
60
80
100
Hours

Figure 1



513

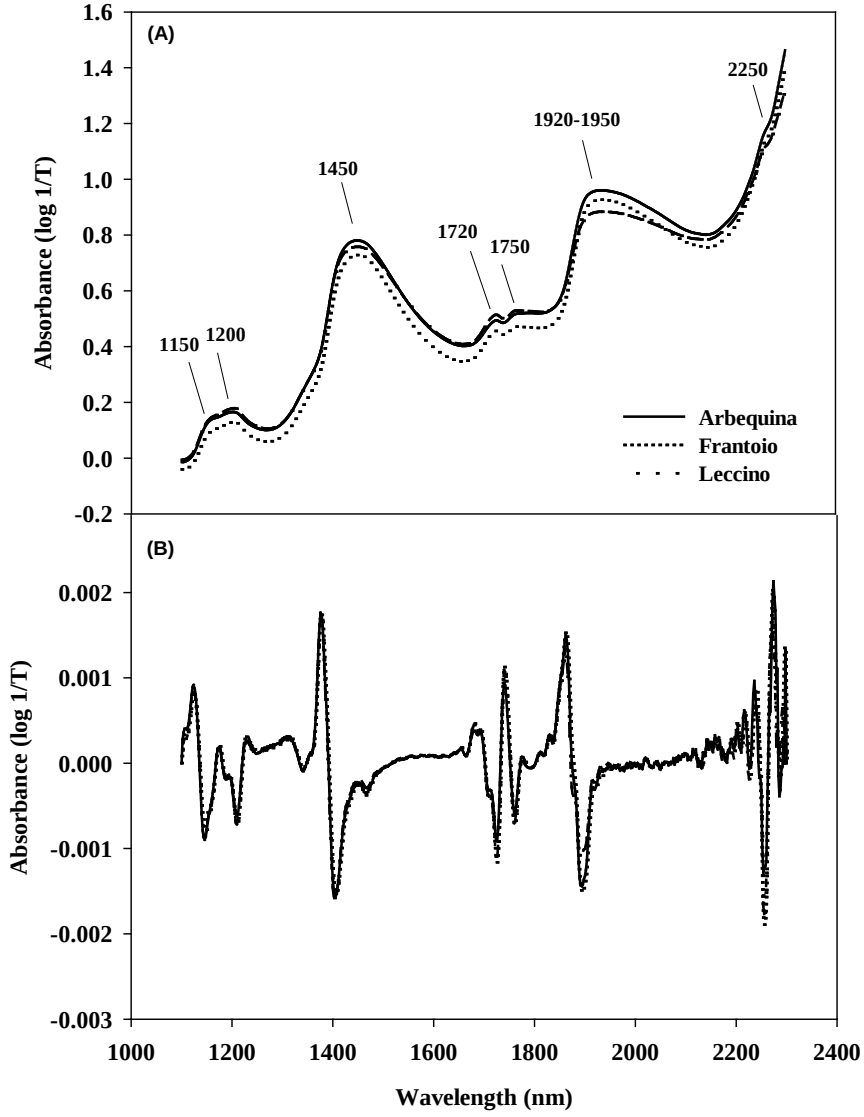
514

515

516

517 **Figure 2**

518



519

520 **Figure 3**

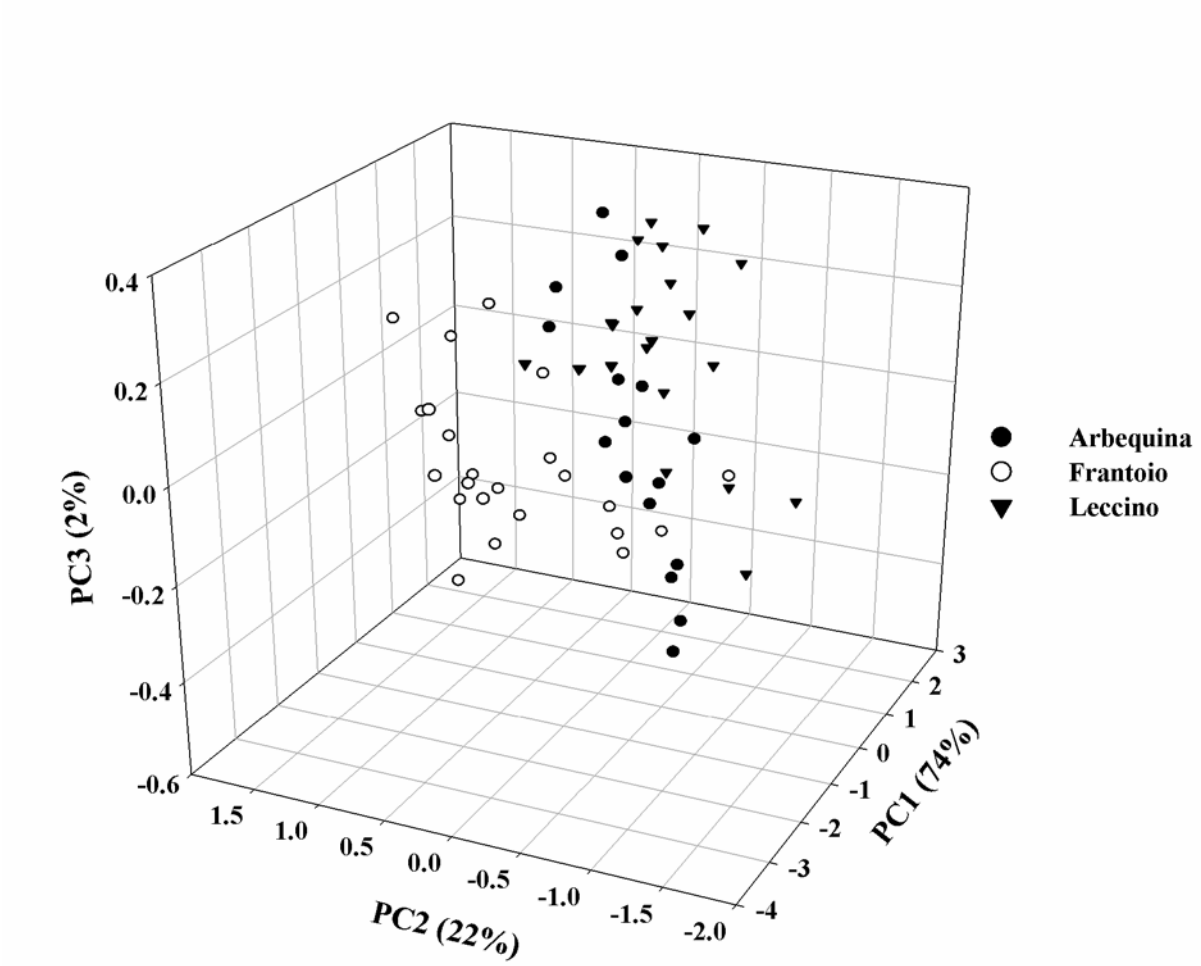

0.4
0.2
0.0
-0.2
-0.4
-0.6
PC3 (2%)
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
PC2 (22%)
3
2
1
0
-1
-2
-3
-4
PC1 (74%)
Arbequina
Frantoio
Leccino

Figure 4

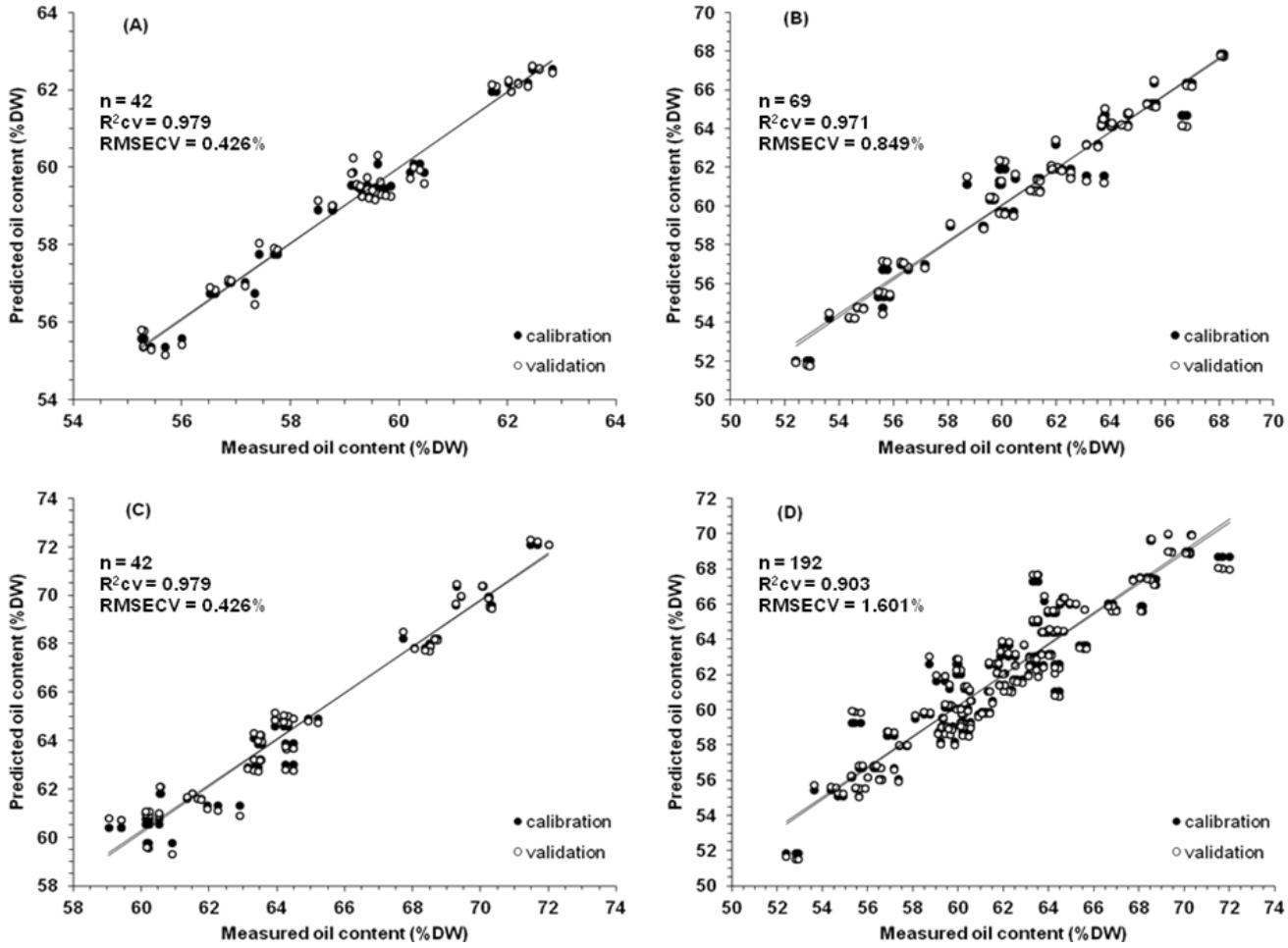

(A)
n = 42
R²cv = 0.979
RMSECV = 0.426%
Predicted oil content (%DW)
Measured oil content (%DW)
calibration
validation
(B)
n = 69
R²cv = 0.971
RMSECV = 0.849%
Predicted oil content (%DW)
Measured oil content (%DW)
calibration
validation
(C)
n = 42
R²cv = 0.979
RMSECV = 0.426%
Predicted oil content (%DW)
Measured oil content (%DW)
calibration
validation
(D)
n = 192
R²cv = 0.903
RMSECV = 1.601%
Predicted oil content (%DW)
Measured oil content (%DW)
calibration
validation

Figure 5